The Effect of Maximal Rate Codes on the Interfering Message Rate

Ronit Bustin*, H. Vincent Poor* and S. Shamai (Shitz)†

* Dept. of Electrical Engineering, Princeton University, email: bustin,poor@princeton.edu

† Dept. of Electrical Engineering, Technion−Israel Institute of Technology,

email: sshlomo@ee.technion.ac.il

Abstract

The effect of "good", point-to-point capacity achieving, code sequences on an additional signal, of bounded variance, transmitted over the additive Gaussian noise channel is examined. For such code sequences, it is shown that their effect, in terms of mutual information, on the additional bounded variance signal, is as if additional additive Gaussian noise has been transmitted. Moreover, the analysis shows that for reliable communication the bounded variance signal must be completely estimated by the receiver ($i.e.$, the minimum mean-square error tends to zero). This result resolves the "Costa Conjecture" regarding the corner points of the two-user Gaussian interference channel for code sequences of bounded variance, and shows that both messages must be reliably decoded.

I. Introduction

A prominent problem in network information theory, open for over 30 years, is the problem of interference. The problem, originated from [1], deals with the full characterization of the capacity region of the interference channel. The interference channel depicts a wide family of communication scenarios, in which several transmitters wish to communicate independent information to their corresponding receivers. Clearly this is an important problem, and as an example one can think of a wireless network in which several end-users share a common space and thus interfere with each other's transmission. When no cooperation is allowed between either the transmitters or the receivers, a single-letter expression for the capacity region of these models is, in general, unknown. The only general result that applies to all possible scenarios is the limiting expression for the capacity region, derived by Ahlswede [2]; however it provides little insight into the actual capacity of a given channel. Two regions for which a single-letter expression for the capacity is available, are the very strong interference regime [3] and the strong interference regime [4]–[6]. In the very strong interference regime, each receiver can completely cancel the interference by exploiting its structure [3]. In the strong interference regime it has been shown that both receivers can decode both

This research was supported in part by the U.S. National Science Foundation under Grant ECCS-1343210, and in part by the Marie Curie International Outgoing Fellowship program under award No. FP7-PEOPLE-IOF-2011-298532. The work of S. Shamai was supported in part by the Israel Science Foundation (ISF) and by the S. AND N. GRAND RESEARCH FUND.

the intended and interfering messages, and thus the capacity region is the same as that of the compound multiple access channel (MAC). A basic information theoretic model to examine the problem of interference is the two-user case, in which two point-to-point links interfere with each other. Information theoretic study of interference channels has been focused mainly on this limited model, with the hope that the insights obtained could be generalized to the general K-user case. Even for this limited model the capacity region is known exactly only for the very strong and strong interference regimes.

The best known achievability scheme for the two-user interference channel is the one presented by Han and Kobayashi (HK) [5]. This scheme achieves capacity for the very strong and strong regimes, and obtains the largest region in all other regimes. Their strategy involves splitting the transmitted information of both users into two parts: private information to be decoded only at the intended receiver and common information that can be decoded at both receivers. They further use joint decoding of the common information from both users and the intended private information. By decoding the common information, part of the interference can be canceled, while the remaining private information from the other user is treated as noise. There are two major drawbacks to the HK scheme. First, the optimal input distributions are unknown. Second, even when a specific distribution is chosen, there are numerous degrees of freedom involved in the calculation of the region, mostly due to the cardinality of the time-sharing parameter. Chong et al. [7] simplified the expression of the HK region ($i.e.$, reduced the number of inequalities and the cardinality of the time-sharing parameter); however it is still prohibitively complex to calculate. Note that the problem becomes even more acute when we generalize to the K-user model, in which the generalization of HK includes for each user a different message for each subset of non-intended receivers [8]. Having said all that, Etkin et al. [9] have shown that in the Gaussian two-user interference channel the HK region is within a single bit per second per hertz from capacity for all values of the channel parameters. Recently the sub-optimality of the HK region has been demonstrated numerically on the clean Z-interference (CZI) channel with binary inputs and outputs [10].

In this work we focus on the important Gaussian two-user interference channel. This channel has attracted considerable attention in recent years (see [11] for an elaborated survey). However, as in general discrete-memoryless channels, apart from strong and very strong interference, all other regimes have no known closed form expression for the capacity region. These other regimes include the weak one-sided interference channel, alternatively known as the Z-interference channel, the weak interference channel and the mixed interference channel. In the Gaussian Z-interference channel only one of the two users interferes with the transmission of the other user, leaving its own channel free of interference. If the interference is strong then the results obtained for the general interference channel apply. On the other hand if the interference is weak, that is, the interference is received with weaker power at the receiver that is interfered with, the capacity region is unknown. In the Gaussian weak interference channel the added interference at each receiver is received with weaker power than at its intended receiver. Finally, the Gaussian mixed interference has one receiver receiving weak interference and the other receiving strong interference.

Since the capacity is generally unknown, our best description of the capacity region is through its inner and outer bounds. The first to provide general outer bounds were Carleial [12] and Sato [13]. In [12] Carleial introduced the

genie-aided approach, that is, outer bounding the capacity region by providing one or both of the receivers additional information regarding the intended or interfering message or both. In the Gaussian interference channel, this side information is usually taken to be some noisy version of some linear combination of the messages (where the noise may be correlated or not with the noise of the channel). In [14] Sato derived a new outer bound specifically for the *degraded* Gaussian interference channel (*degraded* in the sense that the linear combination of the messages received by one of the receivers is a *degraded* version of the combination received by the other receiver). Subsequently, Costa [15] showed that the *degraded* Gaussian interference channel is equivalent to the weak Z-interference channel. Thus, the outer bound derived by Sato provides an outer bound on the capacity region of the weak Gaussian Z-interference channel. This is the best known outer bound for this problem, to date. The derivation of Sato is based on the capacity region of the *degraded* Gaussian broadcast channel (BC). The idea being that the only difference between the two problems is that the power splitting between the two messages sent is strict for the interference channel, while for the BC it is unrestricted. Thus, the capacity region of the *degraded*, BC with total power equal to the sum-power constraint of the two users in the interference channel, is an outer bound on the capacity region of the *degraded* interference channel and the weak Z-interference channel. This outer bound has been rederived in [16], where the authors used the extremal inequality [17] for their derivation, and in [18], where the authors used the entropy power inequality (EPI) [19] for their derivation. In [20] the authors provide a simple derivation of the bound using the relationship between information theory and estimation theory (later referred to as the I-MMSE approach and further detailed). All three derivations relied on Ahlswede's limiting expression for the capacity [2]. Recently, Rioul and Costa [21] targeted specifically the achievable corner point on Sato's outer bound (more about that in the sequel) and provided another elegant derivation which does not rely on Ahlswede's limiting expression [2] but rather on the *almost Gaussian* properties of such a point.

In 2004, Kramer [22] provided two new outer bounds for the two-user Gaussian interference channel. The first outer bound was a genie-aided outer bound. Kramer has shown that this outer bound can also be derived using a different method of linear minimum mean square error estimation. At the same paper, Kramer obtained a second outer bound, which outperforms the first one. This outer bound takes the intersection of the outer bounds on the two underlying Z-interference channels. For these enhanced channels, Kramer uses either Sato's outer bound if the one-sided interference is weak or the capacity region if the one-sided interference is strong. Another highly important work is that of Etkin et al. [9]. In this work the authors derived a new outer bound for the Gaussian interference channel (which we will refer to as the ETW outer bound), using the genie-aided approach. Their motivation was to obtain an outer bound similar in structure to the HK region. Thus, allowing a comparison of the two regions. Using their newly derived outer bound they were able to prove that the HK region is within a single bit per second per hertz (bit/s/hertz) from capacity. This result can also be obtained using the methods of Telatar and Tse in [23]. The result has also been extended to include a common message in [24] and to the multiple-input-multiple-output (MIMO) setting in [25]. Both Kramer's outer bound [22, Theorem 2] and the ETW outer bound improve the outer bounds in [12] and [13]; however no single one is better than the other. The "rule of thumb" is that at low signal-to-noise ratios (SNRs) Kramer's outer bound is tighter and at high SNRs the ETW outer bound

is tighter. Successively, several new outer bounds for the weak interference regime have been derived. All of these bounds are genie-aided bounds. The main differences between them are the freedom they allow in the selection of the genie, the linear combination of rates they bound, and the method used for bounding. These include, Motahari and Khandani's outer bound in [16], the outer bound of Annapureddy and Veeravalli in [18], the outer bound of Shang, Kramer and Chen in [26], the sum-rate outer bound of Etkin in [27] and the outer bound of Chaaban and Sezgin [28]. As noted in [29] the bounds of [16] and [18] are an enhanced version of the ETW bound [29, Theorem 2] (referred to as the "enhanced ETW bound") and contain both the ETW bound and Kramer's bound as special cases. The bound derived in [29, Theorem 4] is shown to be tighter for a certain range of parameters. Thus, the best known outer bound, to date, is the intersection of the "enhanced ETW bound" and [29, Theorem 4].

The capacity region contains a few special boundary points that are important attributes of the region. These include the sum-rate capacity point, which is the point of maximum sum-rate, and the corner points of the capacity region, which are those in which one of the two users transmits at its maximum possible rate. These points have attracted considerable attention. Of course, they are known for the very strong and strong Gaussian interference channel (since the capacity region is fully known). However, they are partially known in many of the other regimes, for which the capacity region is unknown. For the Gaussian Z-interference channel Sato's outer bound [14] contains an achievable corner point when the interferer transmits at maximum rate. This was observed by Sason in [30], who has shown that this corner point is actually the sum-rate capacity point. The weak interference regime is the only regime in which the sum-capacity is *not* a corner point of the capacity region. In this regime the sum-rate capacity is known for a sub-region of parameters, referred to as the "noisy interference" region, in which the sum-rate is achieved by treating interference as noise (TIN). This was revealed, independently, in [16], [26] and [18]. Another result is the sum-rate capacity of the mixed interference channel which was obtained, independently, by Motahari and Khandani in [16] and by Tuninetti and Weng in [31], and is one of the corner points. In [9] is was shown, using the definition of generalized degrees of freedom (GDoF), that in some asymptotic regimes the gap between some lower and upper bound on the sum-rate shrinks to zero, thus obtaining an asymptotic characterization of the sum-capacity. Evidently, not only is the capacity region unknown in general, but also several of its basic attributes, are unknown. A well-known conjecture regarding the missing corner points originated in [15] and [30] and restated in [32, Conjecture 1] (we refer to is as "Costa's Conjecture") is that when one of the two users transmits at maximum point-to-point rate while suffering from interference, the rate of the interference must be limited such that reliable decoding of the interference can be performed while considering the intended transmission as additive Gaussian noise. The achievability of such a corner point is straightforward using Gaussian codebooks: first reliably decode the interference while considering the intended transmission as additional additive Gaussian noise, remove it, and then reliably decode the intended transmission. This conjecture has been investigated in [32] specifically for the weak Gaussian interference channel, where informative bounds have been derived. The bounds are asymptotically tight as the SNR of both transmissions goes to infinity, thus strongly support the conjecture. As noted above the only regime in which the sum-rate is not a corner point of the capacity region is the weak interference regime. As such [32] examines the excess rate, which is the gap between the sum-capacity and the sum-rate at the corner points.

For this quantity [32] performs both an asymptotic analysis, analogous to the GDoF [9], showing the asymptotic tightness of the bounds, and provides improved bounds for finite power. More recently [21] and [33] provide further insights, supporting the "Costa Conjecture". Another important result is the sum-degrees of freedom (DoF) of the K-user Gaussian interference channel for which a general formula has been obtained by Wu et al. [34]. The formula is a single-letter optimization, over the K-dimensional input distribution, of a linear combination of information dimensions [35].

The basic problem when dealing with interference channels is the rivalry between the two users, $i.e.$, the fact that increasing the rate of one user is harmful to the other. Specifically in the Gaussian setting, when looking at a point-to-point channel, it is well know, due to maximum entropy results and the EPI [19], that Gaussian inputs are the best inputs to fight Gaussian noise and Gaussian noise is the worst noise for Gaussian inputs [36], [37]. However, when there are other users in the network, it is not clear that Gaussian is the best input distribution for the overall system. Assuming a two-user interference channel, the first user uses Gaussian inputs, then the second should also use Gaussian inputs to maximize its own rate; however this is the most harmful interference on the first user. This issue is discussed by Abbe and Zheng in [38] (and references therein), where they developed a novel technique to analyze a class of non-Gaussian input distributions over Gaussian noise channels. Using this technique they show that non-Gaussian input distributions outperform Gaussian ones, when limiting the decoder to TIN and the interference is above a certain threshold.

The interference channel is seemingly constructed from two well investigated channels, the BC and the MAC. However, since the receivers are interested only in their intended message and not in the interfering message, the solution cannot be constructed directly from the known capacity region of these two models. In other words, neither channel model considers the effect of a transmission on the unintended receiver. This effect has been the focus of two recent papers. In [39] the authors provide the rate-disturbance region which describes the minimum possible disturbance that can be attained at some interfered-with user for any given rate that can be transmitted reliably to the intended receiver. The disturbance is measured by the mutual information between the codeword and the output at the interfered-with user. A similar question was examined in [40], where the disturbance was measured by the minimum mean-square error (MMSE). In other words the question raised and answered was: given a constraint on the MMSE at some given SNR what is the maximum possible rate of transmission? It was shown that Gaussian superposition codes are optimal in this sense and obtain the maximum possible rate given the MMSE constraint. This provides some engineering support for the good performance of the HK achievability scheme.

Having emphasized the difficulty, a central question dealt with in this context is how to cope with interference. The most basic approach is TIN. This approach has been shown GDoF optimal for the K-user Gaussian interference channel for some range of SNRs [41]. The optimality of TIN in general has been investigated in numerous papers (for example [42] and [43]) and it is well-known that it is sub-optimal in general (see for example [44]). The traditional alternatives to TIN are successive decoding and joint decoding. Specifically in the Gaussian regime these methods were examined in [45] for optimal point-to-point codes and in [46] where the authors examine the interference channel from the point of view of a single transmitter-receiver pair, which is being interfered with.

The authors of [46] proposed a strategy to determine the rate, by disjoining the set of interfering users into two disjoint subsets, namely the set of users where interference is decodable and the set of users where interference is nondecodable. The authors show that, when assuming that interference is Gaussian, their strategy achieves capacity. A more recent approach is that of non-unique decoding (or indirect decoding) as described in [47] and [48]. This approach allows ambiguity in the decoding of unintended messages, as opposed to unique decoding which requires the reliable decoding of the entire message subset. This approach has been shown to achieve capacity of interference channels under the restriction to random codebooks, superposition and time-sharing [49] and also the optimality of HK for the two-user interference channel has been shown under these restrictions. This optimality has recently been shown also through exact expressions for the error exponents assuming such code sequences [50] and indicates that rate-wise the HK scheme cannot be improved upon by merely using the maximum-likelihood decoder. However, it was recently shown that similar results can be attained by unique decoding (in all settings that used non-unique decoding), and thus it is not yet clear whether non-unique decoding can enlarge the achievable rate region [51]. Specifically, when considering the additive white Gaussian noise (AWGN) channel an approach related to non-unique decoding is that of assuming some structure on the noise. In [52], the authors examined alternatives to TIN, assuming the receiver knows the constellation set used by the interferer. This makes the interference plus noise a mixed Gaussian process. Under these assumptions the authors develop an achievable rate, with improved sum-rate as compared to that obtained when using Gaussian codebooks and TIN. More recently it was shown that assuming mixed Gaussian noise is sum-GDoF optimal for the symmetric interference channel [53]. Due to the complexity of the problem some interesting results were obtained by placing assumptions either on the codebook structure, the encoders or the receivers. This was done by Bandemer et al. [49] as mentioned above. Baccelli et al. [45] establish the capacity region of the K-user Gaussian interference channel, when all users are constrained to use point-to-point codes. The capacity is shown to be strictly larger, in general, than the achievable rate regions when using TIN, using successive interference cancelation decoding, and using joint decoding. More recently the problem of an oblivious receiver that lacks the knowledge of the codebooks has been investigated in [54] and some capacity results have been obtained.

In this paper we focus on a complementary problem to the one considered by Baccelli et al. [45]. We also assume the use of a “good”, capacity achieving code sequence; however we assume that it is used only for the transmission of the interfered-with user. The interference may be any general signal of bounded variance (not necessarily a code sequence). Given these assumption the main question discussed in this work is: What are the limitations that the transmission of a “good” code sequence inflicts on the properties of the interfering signal (in terms of input-output mutual information)? This question allows us to answer a more central question which is: What is the maximum rate of the interfering message allowing reliable communication of the interfered-with transmission of maximum rate ($i.e.$ transmission of a “good” code sequence)? Our answer to the first question is that for the relevant set of SNRs the effect of such a transmission is as if additive Gaussian noise has been transmitted over the channel. Thus, the answer to the second question is that the MMSE of the interference must go to zero at the receiver, meaning we must be able to estimate the interference with vanishing error as well as reliably decode the intended message

(thus answering [32, Question 2]). These results provide the missing corner points of reliable communication in the two-user Gaussian interference channel, resolving the "Costa Conjecture" for bounded variance inputs, [15] and [32, Conjecture 1].

More specifically, the focus of this work is on the following discrete-time, memoryless channel:

$$\boldsymbol{Y}_n = \sqrt{\mathsf{snr}_1}\boldsymbol{X}_n + \sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n + \boldsymbol{N}_n \tag{1}$$

where $\boldsymbol{N}_n$ represents a standard additive Gaussian noise vector with independent components, $\boldsymbol{X}_n$ carries the intended message and $\boldsymbol{Z}_n$ is the interfering signal. $\boldsymbol{X}_n$ and $\boldsymbol{Z}_n$ are independent of each other (no cooperation between the transmitters) and independent of the additive Gaussian noise vector. We further assume, for simplicity and without loss of generality, that both inputs are zero-mean. The subscript n denotes that all vectors are length n vectors. snr_1 and snr_2 are both non-negative scalar parameters and allow us to assume an average power constraint of 1 on both $\boldsymbol{X}_n$ and $\boldsymbol{Z}_n$ without loss of generality. The parameter a is also a non-negative scalar parameter, and is used here for consistency with the two-user Gaussian interference problem discussed in Section VI. We assume that there is a sequence of point-to-point capacity achieving codebooks ($i.e.$, that approach capacity, as $n \to \infty$, with vanishing probability of error). $\boldsymbol{X}_n$ carries a message from the length n codebook. Thus, when $n \to \infty$,

$$\mathsf{R}_x = \frac{1}{2}\log\left(1 + \mathsf{snr}_1\right). \tag{2}$$

As stated above, the main question is: What limitations does this inflict on the interfering signal, $\{\boldsymbol{Z}_n\}_{n\geq 1}$, in terms of input-output mutual information?

The approach taken in this work is the exploration of the behavior of the mutual information quantities as a function of the SNR. This provides us with the mutual information value at a set of different outputs that are seemingly unrelated to the given problem, as opposed to the value of the mutual information quantities only at the relevant outputs. For this approach we employ the fundamental relationship between information theory and estimation theory shown in [55], also known as the I-MMSE relationship. This relationship fits our approach since for any input going through the AWGN channel it relates the derivative of the input-output mutual information with the MMSE function when estimating the input from the output. This allows us to draw important conclusions from the analysis of the optimal estimators and their MMSE function. Note that the I-MMSE relationship holds for an input-output through the AWGN channel; however not all quantities fit this relation. A central ingredient in our analysis is the assumption of a "good" code sequence used by one of the two transmitters, meaning a code sequence that attains point-to-point capacity over the un-interfered-with AWGN channel (no additional transmission over the channel). The significance of this assumption is that the exact behavior of such codes over the AWGN channel in terms of the mutual information (and MMSE) is known for all SNRs [56]–[58]. These papers have shown that the mutual information of "good" code sequences is as if an independent and identically distributed (i.i.d.) Gaussian input has been transmitted, up to the SNR of reliable decoding. In some sense the results presented here are an extension of this observation, as they show that "good" code sequences also affect an additional transmission through the channel as if they were i.i.d. Gaussian noise (in terms of the mutual information).

The rest of this paper is organized as follows: Section II contains some preliminary results. We then revisit the properties of “good”, point-to-point capacity achieving code sequences in Section III. We formulate the problem in Section IV and then give our main results in Section V. These results are then applied to the open problem of the corner points of the capacity region of the two-user Gaussian interference channel in Section VI. Finally, we conclude the paper in Section VII.

II. Preliminary Definitions and Results

A. Preliminary Definitions

We consider the scalar setting, in which length n codewords are to be transmitted over the channel (n subsequent uses of the channel). We then consider the behavior of such transmissions in the limit as $n \to \infty$, so as to conclude regarding the maximum possible rates given our assumptions on the inputs. As such we denote any vector of finite length n with a subscript n, and when $n \to \infty$ we remove the subscript. More specifically, assuming a length n input to a channel, denoted by $\boldsymbol{X}_n$, and the corresponding output, denoted by $\boldsymbol{Y}_n$, the per-component mutual information is

$$\frac{1}{n} I\left(\boldsymbol{X}_n; \boldsymbol{Y}_n\right). \tag{3}$$

Given a random process over $\{\boldsymbol{X}_n\}_{n\geq 1}$, if the limit of (3) exists as $n \to \infty$, it will be denoted as

$$I\left(\boldsymbol{X}; \boldsymbol{Y}\right) = \lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \boldsymbol{Y}_n\right). \tag{4}$$

A joint input-output distribution for which the above holds is known as *information stable* [59]. In this work we assume only *information stable* input-output distributions. The MMSE matrix when estimating $\boldsymbol{X}_n$ from the channel output $\boldsymbol{Y}_n$ is defined as follows:

$$\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n) = \mathsf{E}\left\{\left(\boldsymbol{X}_n - \mathsf{E}\left\{\boldsymbol{X}_n|\boldsymbol{Y}_n\right\}\right)\left(\boldsymbol{X}_n - \mathsf{E}\left\{\boldsymbol{X}_n|\boldsymbol{Y}_n\right\}\right)^{\mathsf{T}}\right\}. \tag{5}$$

The MMSE function is then defined as

$$\mathsf{MMSE}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = \frac{1}{n}\mathsf{Tr}(\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n)). \tag{6}$$

If these quantities converge as $n \to \infty$ we denote the limits as

$$\mathsf{MMSE}(\boldsymbol{X}|\boldsymbol{Y}) = \lim_{n\to\infty} \mathsf{MMSE}(\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma)). \tag{7}$$

and

$$\mathbf{E}_{\boldsymbol{X}}(\boldsymbol{Y}) = \lim_{n\to\infty} \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n). \tag{8}$$

When this is not the case ($i.e.$, the MMSE function and/or the MMSE matrix do not converge) we will consider the $\limsup$s of these sequences.

B. The I-MMSE Relationship

The main approach used here is the I-MMSE approach, that is to say that we make use of the fundamental relationship between the mutual information and the MMSE in the Gaussian channel and its generalizations [55], [60]–[62]. Even though we are examining scalar settings, the n-dimensional version of this relationship is required since we are looking at the transmission of length n codewords through the channel. In our setting the relationship is as follows:

$$\frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{MMSE}(\boldsymbol{X}_n | \sqrt{\gamma}\boldsymbol{X}_n + \boldsymbol{N}_n) \mathsf{d}\gamma \tag{9}$$

where $\boldsymbol{N}_n$ is standard additive Gaussian noise, and $\boldsymbol{X}_n$ is the input signal of any arbitrary distribution (as long as the above mutual information is finite [63]).

We now take the limit as $n \to \infty$ on both sides of the I-MMSE relationship (9). For random processes for which the relevant MMSE quantities converge as $n \to \infty$, the exchange of limit and integration is according to Lebesgue's dominated convergence theorem [64], since the MMSE quantities are always bounded. Thus, we have

$$\begin{aligned} I\left(\boldsymbol{X}; \sqrt{\mathsf{snr}}\boldsymbol{X} + \boldsymbol{N}\right) &\equiv \lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right) \\ &= \lim_{n\to\infty} \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{MMSE}(\boldsymbol{X}_n | \sqrt{\gamma}\boldsymbol{X}_n + \boldsymbol{N}_n) \mathsf{d}\gamma \\ &= \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{MMSE}(\boldsymbol{X} | \sqrt{\gamma}\boldsymbol{X} + \boldsymbol{N}) \mathsf{d}\gamma. \end{aligned} \tag{10}$$

When this is not the case ($i.e.$, the MMSE does not converge) we can apply the reverse Fatou's Lemma [64] and conclude that

$$\lim_{n\to\infty} \int_0^{\mathsf{snr}} \mathsf{MMSE}(\boldsymbol{X}_n | \sqrt{\gamma}\boldsymbol{X}_n + \boldsymbol{N}_n) \mathsf{d}\gamma = \int_0^{\mathsf{snr}} \mathsf{MMSE}(\boldsymbol{X} | \sqrt{\gamma}\boldsymbol{X} + \boldsymbol{N})_{sup} \mathsf{d}\gamma \tag{11}$$

where

$$\mathsf{MMSE}(\boldsymbol{X} | \sqrt{\gamma}\boldsymbol{X} + \boldsymbol{N})_{sup} = \limsup_{n\to\infty} \mathsf{MMSE}(\boldsymbol{X}_n | \sqrt{\gamma}\boldsymbol{X}_n + \boldsymbol{N}_n) \tag{12}$$

due to our assumption of *information stable* input-output distributions. This is shown precisely in [65, Appendix A]. In this case the MMSE quantities throughout the paper are the $\limsup$s of the MMSE sequences, which always exist and, since the MMSE function is bounded, are also finite.

C. Uniform Integrability and The Vitali Convergence Theorem

As we examine the behavior of the quantities in the limit, as $n \to \infty$, we need to clarify several notions of convergence and provide some relevant results. The subsequent results best explain the assumptions set on the input distributions in our main results given in Section V. The following results concern the concept of *uniform integrability* [66]. Uniformly integrable families are those in which the contribution of the tail to the expectation is controlled uniformly over all elements. In integrable functions this simply reduces to a small contribution of the tail to the expectation. The importance of this concept in this work is that it allows us to use the "Vitali

Convergence Theorem" [67], [68], a relaxation of the dominated convergence theorem that reduces the requirement of the dominating integrable function to that of a uniformly integrable family.

We begin by giving the definition of *uniform integrability*

Definition 1 (Definition 12.1 [66]): A non-empty family $\mathcal{X}$ of random variables is said to be *uniformly integrable* if

$$\lim_{K \to \infty} \left(\sup_{X \in \mathcal{X}} \mathsf{E}\left\{ |X| 1_{|X|>K} \right\} \right) = 0. \tag{13}$$

Next we give two simple conditions for *uniform integrability*.

Lemma 1 (Corollary 12.8 [66]): For $p > 1$, let $\mathcal{X}$ be a nonempty family of random variables bounded in $\mathcal{L}^p$, $i.e.$, such that $\sup_{X \in \mathcal{X}} \parallel X \parallel_{\mathcal{L}^p} < \infty$. Then $\mathcal{X}$ is uniformly integrable.
It is well-known that $\mathcal{L}^1$ boundedness does not suffice for *uniform integrability* (see [66, Remark 12.4]). The second condition for *uniform integrability* is the following:

Lemma 2 (Proposition 1.10 [69]): Suppose that $\{X_n\}_{n \geq 1}$ is a nonempty family of random variables bounded in $\mathcal{L}^1$, and X_n converges to X *a.s.*. Then X_n converges to X in $\mathcal{L}^1$ if and only if $\{X_n\}_{n \geq 1}$ is *uniformly integrable*.

Next we give the "Vitali Convergence Theorem":

Theorem 1 ([67]): Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a positive measure space. If

1) $\mu(\mathcal{X}) < \infty$
2) $\{f_n\}$ is *uniformly integrable*
3) $f_n(X) \to f(X)$ *a.s.* as $n \to \infty$ and
4) $|f(X)| < \infty$ *a.s.*

then the following hold:

1) $f \in \mathcal{L}^1(\mu)$
2) $\lim_{n \to \infty} \int_{\mathcal{X}} |f_n - f| \mathsf{d}\mu = 0$.

As stated in [70, Section 3] and in [71, Proposition 11], convergence theorems can be extended to the conditional expectation, meaning that they hold for any σ-algebra contained in $\mathcal{F}$ *a.s.*.

III. On "Good" Point-to-Point Code Sequences

In this section we revisit "good", capacity achieving code sequences over the point-to-point AWGN channel. Assume a "good" code sequence which attains reliable communication over the AWGN channel at snr, under a power constraint of 1, meaning that all codewords in such a sequence comply with

$$\frac{1}{n} \parallel \boldsymbol{x}_n \parallel \leq 1, \quad \forall \boldsymbol{x}_n \in \mathsf{C}_n \text{ and } \forall n \geq 1. \tag{14}$$

The random process defining such code sequences will have the following per-component mutual information in the limit, as $n \to \infty$:

$$I\left(\boldsymbol{X}; \sqrt{\mathsf{snr}}\boldsymbol{X} + \boldsymbol{N}\right) = \frac{1}{2}\mathsf{log}(1 + \mathsf{snr}). \tag{15}$$

The behavior of this quantity as a function of SNR in the limit, as $n \rightarrow \infty$, has been investigated in [56] and [57] and more recently discussed also in [58] from an I-MMSE perspective. The following result was derived:

Theorem 2 ([56], [57]): Any optimal point-to-point codebook sequence, complying with the power constraint (14), which attains capacity at the output $\boldsymbol{Y}(\mathsf{snr})$ (designed for reliable decoding at snr), has the following mutual information behavior:

$$I\left(\boldsymbol{X}; \sqrt{\gamma \mathsf{snr}}\boldsymbol{X} + \boldsymbol{N}\right) = \begin{cases} \frac{1}{2}\log\left(1+\gamma\mathsf{snr}\right) & \forall \gamma \in [0,1] \\ \frac{1}{2}\log\left(1+\mathsf{snr}\right) & \forall \gamma > 1 \end{cases} \tag{16}$$

where $\boldsymbol{N}$ is a standard additive Gaussian noise vector with independent components. The MMSE behavior in the limit (or $\limsup \mathsf{MMSE}$) follows immediately from the I-MMSE relationship (10) - (11).

Proof: We provide the I-MMSE perspective proof given in [58] in Appendix A for completeness. ■

In words, for any code of maximum possible rate over the Gaussian interference channel, its mutual-information behavior over a clean AWGN channel is known for every γ. As stated in the above theorem, using the I-MMSE relationship [55] in the limit as $n \rightarrow \infty$ (considering the $\limsup$ as explained in Section II-B) one can also conclude from the above that

$$\limsup_{n\rightarrow\infty} \mathsf{MMSE}(\boldsymbol{X}_n | \sqrt{\gamma\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n) = \mathsf{MMSE}(\boldsymbol{X}|\sqrt{\gamma\mathsf{snr}}\boldsymbol{X} + \boldsymbol{N}) = \frac{1}{1+\gamma\mathsf{snr}}, \quad \forall \gamma \in [0,1) \tag{17}$$

and zero for all $\gamma \geq 1$, due to the reliable decoding of the message.

The next claim extends upon Theorem 2 by showing additional properties of "good" code sequences for the Gaussian point-to-point channel. From considering the input-output mutual information and MMSE function in Theorem 2 the next claim regards the MMSE matrix of such codes and also concludes regarding the convergence of the optimal estimator.

Theorem 3: Given a random process over $\{\boldsymbol{X}_n\}_{n\geq 1}$ resulting in a "good" code sequence with reliable decoding from an output of an AWGN channel at snr, we have that

$$\begin{aligned} &\lim_{n\rightarrow\infty} \left(\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n^c) - \mathbf{E}_{\boldsymbol{X}}^{OBL}(\boldsymbol{Y}_n^c)\right) = 0 \\ &\lim_{n\rightarrow\infty} \hat{X}(\boldsymbol{Y}_n^c) = \lim_{n\rightarrow\infty} \hat{X}^{OBL}(\boldsymbol{Y}_n^c) = \frac{\sqrt{\gamma\mathsf{snr}}}{1+\gamma\mathsf{snr}}\boldsymbol{Y}^c \quad a.s. \end{aligned} \tag{18}$$

where $\boldsymbol{N}_n$ is standard additive Gaussian noise, $\boldsymbol{Y}_n^c = \sqrt{\gamma\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n$, $\hat{X}^{OBL}(\boldsymbol{Y}_n^c)$ denoted the optimal bit-wise linear estimator and $\mathbf{E}_{\boldsymbol{X_n}}^{OBL}(\boldsymbol{Y}_n^c)$ denotes the MSE matrix of the optimal bit-wise linear estimator.

Note that we do not claim for convergence of either the MMSE matrix not the MSE matrix of the optimal bit-wise linear estimator but rather only to the convergence of the difference between these two matrices.

Proof: The proof is given in Appendix B. ■

The properties of "good" code sequences have been considered by Shamai and Verdú in [72], where they have shown that the k^{th} order empirical distribution of such codes (when k is of the order of $\mathcal{O}(\log n)$) converges to the mutual information maximizing distribution. In the AWGN case they have shown convergence is in distribution (weak convergence) to the standard i.i.d. Gaussian distribution. A simple conclusion from this which we require in the sequel is given in the next lemma.

Lemma 3: Given that the random process over $\{\boldsymbol{X}_n\}_{n\geq 1}$ is of bounded variance and results in a "good" code sequence with reliable decoding from an output of an AWGN channel at snr, its k^{th} order moments converge point-wise to those of the standard i.i.d. Gaussian distribution. Specifically, taking $k=2$ we have that

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{Tr}\left(\mathbf{R}_{\boldsymbol{X}_n}\right) = 1. \tag{19}$$

Proof: This is a direct result of the fact that weak convergence and uniform integrability guarantee the convergence of the moments [73, Chapter 6]. From [72] we have weak convergence and since the process has bounded variance it is *uniformly integrable* (see Lemma 1). ∎

IV. Problem Formulation

As stated in the Introduction our focus is on the channel model given in (1). To simplify our notation in the sequel we define the following:

$$\boldsymbol{W}_n = \sqrt{\mathsf{snr}_1}\boldsymbol{X}_n + \sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n. \tag{20}$$

We also define

$$\boldsymbol{Y}_n(\gamma) = \sqrt{\gamma}\boldsymbol{W}_n + \boldsymbol{N}_n, \ \text{for } \gamma \geq 0. \tag{21}$$

As we aim to obtain insight regarding the maximum possible rate, we allow the blocklength of the transmitted message to go to infinity, meaning $n \to \infty$. For this purpose we formulate the problem on the set of joint random processes over $\{\boldsymbol{X}_n, \boldsymbol{Z}_n\}_{n\geq 1}$, where the random process over $\{\boldsymbol{Z}_n\}_{n\geq 1}$ is independent of the random process over $\{\boldsymbol{X}_n\}_{n\geq 1}$. We further assume that both random processes have bounded variances.

We further define, for any arbitrary component in the length n vector, *e.g.*, X (a component of $\boldsymbol{X}_n$) and Z (a component of $\boldsymbol{Z}_n$), the following two sequences of random variables:

$$\begin{aligned} G_n &= \mathsf{E}\left\{X|\boldsymbol{Y}_n, \boldsymbol{Z}_n\right\} \\ H_n &= \mathsf{E}\left\{Z|\boldsymbol{Y}_n\right\} \end{aligned} \tag{22}$$

for every n. Note: we chose not to denote the components using a subscript i, $e.g.$ X_i, so as to reduce the complexity of the notation. The observations are identical for every component of the length n vectors. Note that the sequence is well-defined since, for every n, the optimal estimator is known *a.s.*.

A simple consequence regarding the sequences of optimal estimators is given in the following lemma.

Lemma 4: The sequence $\{G_n\}_{n\geq 1}$ and the sequence $\{H_n\}_{n\geq 1}$ are both *uniformly integrable*.

Proof: By showing that both sequences are $\mathcal{L}^2$ bounded we can conclude that due to Lemma 1 the sequences are *uniformly integrable*. We show this property only for the sequence $\{G_n\}_{n\geq 1}$, as the same approach can be used on the sequence $\{H_n\}_{n\geq 1}$. As these are sequences of optimal MMSE-wise estimators we have that

$$\begin{aligned} |\mathsf{E}\left\{G_n^2\right\}| &= |\mathsf{E}\left\{X^2\right\} - \mathsf{MMSE}(X|\boldsymbol{Y}_n, \boldsymbol{Z}_n)| \\ &\leq |\mathsf{E}\left\{X^2\right\}| + |\mathsf{MMSE}(X|\boldsymbol{Y}_n, \boldsymbol{Z}_n)| \\ &\leq |\mathsf{E}\left\{X^2\right\}| + \mathsf{Var}(X) < \infty \end{aligned} \tag{23}$$

where the last transition is due to the assumption that $\{\boldsymbol{X}_n\}_{n\geq 1}$ has bounded variance. This concludes the proof. ∎

The central assumption in this work is that $\{\boldsymbol{X}_n\}_{n\geq 1}$ defines a code sequence such that

$$\mathsf{R}_x = I\left(\boldsymbol{X};\boldsymbol{Y}(1)\right) = \frac{1}{2}\log\left(1+\mathsf{snr}_1\right). \tag{24}$$

Since

$$I\left(\boldsymbol{X};\sqrt{\mathsf{snr}_1}\boldsymbol{X}+\boldsymbol{N}\right) \geq I\left(\boldsymbol{X};\boldsymbol{Y}(1)\right) \tag{25}$$

where $\boldsymbol{N}$ is a standard additive Gaussian noise vector with independent components, and due to the maximum entropy result we can conclude that

$$I\left(\boldsymbol{X};\sqrt{\mathsf{snr}_1}\boldsymbol{X}+\boldsymbol{N}\right) = \frac{1}{2}\log\left(1+\mathsf{snr}_1\right). \tag{26}$$

As stated in Theorem 2 (see also [56], [57] and [58]) we can further conclude that

$$I\left(\boldsymbol{X};\sqrt{\gamma\mathsf{snr}_1}\boldsymbol{X}+\boldsymbol{N}\right) = \frac{1}{2}\log\left(1+\gamma\mathsf{snr}_1\right) \quad \forall\gamma\in[0,1] \tag{27}$$

and use Theorem 3 to conclude the *a.s.* convergence of the optimal estimator to the optimal bit-wise linear estimator, that is

$$\lim_{n\to\infty} G_n = \frac{\sqrt{\gamma\mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}\left(Y(\gamma)-\sqrt{\gamma a\mathsf{snr}_2}Z\right) \quad \textit{a.s.}. \tag{28}$$

We now wish to make two observations. First, for all $\gamma\in[0,\infty)$

$$I\left(\boldsymbol{W};\boldsymbol{Y}(\gamma)\right) = I\left(\boldsymbol{X},\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right). \tag{29}$$

This can be seen using the differential entropy function, denoted here by $h(\cdot)$, as follows:

$$\begin{aligned} I\left(\boldsymbol{W};\boldsymbol{Y}(\gamma)\right) &= h(\boldsymbol{Y}(\gamma)) - h(\boldsymbol{Y}(\gamma)|\boldsymbol{W}) \\ &= h(\boldsymbol{Y}(\gamma)) - h(\boldsymbol{Y}(\gamma)|\boldsymbol{X},\boldsymbol{Z}) = I\left(\boldsymbol{X},\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right). \end{aligned} \tag{30}$$

Second, for all $\gamma\in[0,1]$,

$$\begin{aligned} I\left(\boldsymbol{W};\boldsymbol{Y}(\gamma)\right) &= I\left(\boldsymbol{X},\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) \\ &= I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) + I\left(\boldsymbol{X};\boldsymbol{Y}(\gamma)|\boldsymbol{Z}\right) \\ &= I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) + I\left(\boldsymbol{X};\sqrt{\gamma\mathsf{snr}_1}\boldsymbol{X}+\boldsymbol{N}\right) \\ &= I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) + \frac{1}{2}\log\left(1+\gamma\mathsf{snr}_1\right) \end{aligned} \tag{31}$$

where we make another use of our "good" code assumption over $\{\boldsymbol{X}_n\}_{n\geq 1}$. Taking the derivative with respect to γ, and using the I-MMSE relationship [55], we have for $\gamma\in[0,1)$

$$\begin{aligned} \mathsf{MMSE}(\boldsymbol{W}|\boldsymbol{Y}(\gamma)) &= 2\frac{\mathsf{d}}{\mathsf{d}\gamma}I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) + \mathsf{MMSE}(\boldsymbol{X}|\sqrt{\gamma\mathsf{snr}_1}\boldsymbol{X}+\boldsymbol{N})\cdot\mathsf{snr}_1 \\ &= 2\frac{\mathsf{d}}{\mathsf{d}\gamma}I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) + \frac{\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1} \end{aligned} \tag{32}$$

where the last transition is due to (17). Recall that the above MMSE functions are the $\limsup$s of the MMSE sequences, which do not necessarily converge point-wise. Due to reliable decoding at $\gamma = 1$ we have for $\gamma \geq 1$

$$\mathsf{MMSE}(\boldsymbol{W}|\boldsymbol{Y}(\gamma)) = 2\frac{\mathsf{d}}{\mathsf{d}\gamma} I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right). \tag{33}$$

Note that $I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right)$ is not the transmission of a codeword over an AWGN channel, and thus we do not know much about its derivative. Our main result targets exactly the behavior of this quantity, showing that under our assumptions it behaves as if $\boldsymbol{Z}$ were transmitted over an AWGN channel.

V. Main Result

Our main result is the following:

Theorem 4: For any independent random process over $\{\boldsymbol{X}_n, \boldsymbol{Z}_n\}_{n\geq 1}$, both of bounded variances, where $\{\boldsymbol{X}_n\}_{n\geq 1}$ results in a “good” code sequence with reliable decoding from an output of an AWGN channel at snr_1, we have that

$$2\frac{\mathsf{d}}{\mathsf{d}\gamma} I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) = \begin{cases} \mathsf{MMSE}\left(\boldsymbol{Z}|\sqrt{\frac{\gamma a\mathsf{snr}_2}{1+\gamma\mathsf{snr}_1}}\boldsymbol{Z}+\boldsymbol{N}\right)\cdot\frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2}, & \gamma\in[0,1) \\ \mathsf{MMSE}\left(\boldsymbol{W}|\boldsymbol{Y}(\gamma)\right), & \gamma\geq 1 \end{cases}. \tag{34}$$

This result can be viewed as an extension of the result in [56] and [57] as it shows that a “good” (optimal point-to-point) code sequence not only behaves as an i.i.d. Gaussian random vector when we examine its input-output mutual information and MMSE over the AWGN channel, but also has an i.i.d. Gaussian effect on the additional interfering input, $\boldsymbol{Z}_n$. In other words, as $n\to\infty$, a “good” code sequence can be regarded as additional additive i.i.d. Gaussian noise, when the mutual information and MMSE of $\boldsymbol{Z}_n$ and the output are considered.

The main result required in order to prove Theorem 4 is the following:

Theorem 5: The following behavior holds for any independent random process over $\{\boldsymbol{X}_n, \boldsymbol{Z}_n\}_{n\geq 1}$, both of bounded variances, such that $\{\boldsymbol{X}_n\}_{n\geq 1}$ results in a “good” code sequence with reliable decoding from an output of an AWGN channel at snr_1:

$$2\frac{\mathsf{d}}{\mathsf{d}\gamma} I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) = \mathsf{MMSE}\left(\boldsymbol{Z}|\boldsymbol{Y}(\gamma)\right)\cdot\frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2}.$$

Proof: We begin by examining the following optimal finite dimensional MMSE-wise estimator:

$$\begin{aligned}
\mathsf{E}\left\{\boldsymbol{W}_n|\boldsymbol{Y}_n(\gamma)\right\} &= \mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n+\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\} \\
&= \mathsf{E}\left\{\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n+\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma),\boldsymbol{Z}_n\right\}|\boldsymbol{Y}_n(\gamma)\right\} \\
&= \mathsf{E}\left\{\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma),\boldsymbol{Z}_n\right\}+\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\} \\
&= \mathsf{E}\left\{\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma),\boldsymbol{Z}_n\right\}|\boldsymbol{Y}_n(\gamma)\right\}+\mathsf{E}\left\{\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}.
\end{aligned} \tag{35}$$

Now, we consider $\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma),\boldsymbol{Z}_n\right\}$:

$$\begin{aligned}
\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma),\boldsymbol{Z}_n\right\} &= \mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma)-\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n,\boldsymbol{Z}_n\right\} \\
&= \mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma)-\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n\right\}
\end{aligned} \tag{36}$$

where the second equality is due to the fact that $\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z}_n = \sqrt{\gamma \mathsf{snr}_1}\boldsymbol{X}_n + \boldsymbol{N}_n$ is independent of $\boldsymbol{Z}_n$. The estimator $\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n | \boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\}$ is simply the estimator of $\boldsymbol{X}_n$ from a clean AWGN channel. As shown in Section III the above sequence of optimal estimators converges *a.s.* to the optimal bit-wise linear estimator due to Theorem 3, that is,

$$\begin{aligned}\lim_{n\to\infty} \mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n | \boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} &= \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z}) \\ \lim_{n\to\infty} G_n &= \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}Z). \end{aligned} \tag{37}$$

Note that the above is a well-defined sequence of random variables since the optimal MMSE estimator is *a.s.* unique for every n. Moreover, for every finite n the optimal estimator is markedly different from the optimal bit-wise linear estimator and only as $n \to \infty$ does it converge to it. Furthermore, as shown in Lemma 4 the sequence $\{G_n\}_{n\geq 1}$ is *uniformly integrable*. Thus, we may apply the "Vitali Convergence Theorem" (see Theorem 1) to conclude that for every $\sigma-$algebra $\mathcal{G}$

$$\begin{aligned}\lim_{n\to\infty} \mathsf{E}\left\{G_n | \mathcal{G}\right\} &= \mathsf{E}\left\{G | \mathcal{G}\right\} \quad \textit{a.s.} \\ \lim_{n\to\infty} \mathsf{E}\left\{\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n | \boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} | \mathcal{G}\right\} &= \mathsf{E}\left\{\left.\frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z})\right| \mathcal{G}\right\} \quad \textit{a.s.}. \end{aligned} \tag{38}$$

We now define the following linear estimator of X from $\boldsymbol{Y}_n$ and $\boldsymbol{Z}_n$:

$$\hat{X}^{SBL} = \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\left(Y - \sqrt{\gamma a \mathsf{snr}_2}Z\right). \tag{39}$$

These are bit-wise linear estimators, however they are not necessarily the optimal bit-wise linear estimators since the variance of $\{\boldsymbol{X}_n\}_{n\geq 1}$ may vary. Furthermore, the above estimators are set for all n (taking the $\lim$ as $n \to \infty$ has no effect). Since this is an integrable function (which trivially bounds the sequence for all n) we may apply the Lebesgue dominated convergence theorem [64] (conditioned on a σ-algebra $\mathcal{G}$) to conclude that

$$\begin{aligned}\lim_{n\to\infty} \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z}_n) &= \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z}) \\ \lim_{n\to\infty} \mathsf{E}\left\{\left.\frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z}_n)\right| \mathcal{G}\right\} &= \mathsf{E}\left\{\left.\frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z})\right| \mathcal{G}\right\} \quad \textit{a.s.}. \end{aligned} \tag{40}$$

Note also that, as shown in Theorem 3, since we are considering a "good" code sequence, both the optimal estimator and the optimal bit-wise linear estimator converge *a.s.* to these sub-optimal estimators in the limit as $n \to \infty$. More importantly, the performance of these sub-optimal bit-wise linear estimators can be easily derived (see Corollary 3). Putting both observations together we can conclude that for any $\sigma-$algebra $\mathcal{G}$

$$\lim_{n\to\infty} \mathsf{E}\left\{\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n | \boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} | \mathcal{G}\right\} = \lim_{n\to\infty} \mathsf{E}\left\{\left.\frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z}_n)\right| \mathcal{G}\right\} \tag{41}$$

and specifically,

$$\lim_{n\to\infty} \mathsf{E}\left\{\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n | \boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} | \boldsymbol{Y}_n\right\} = \lim_{n\to\infty} \mathsf{E}\left\{\left.\frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a \mathsf{snr}_2}\boldsymbol{Z}_n)\right| \boldsymbol{Y}_n\right\} \tag{42}$$

a.s..

Before using the above equality in the estimator of $\boldsymbol{W}_n$ (35) we need to consider what happens in the limit. Note that we did not assume that the sequence of optimal estimators of $\boldsymbol{W}_n$ converges as $n \to \infty$ in any sense. Moreover, we do not even know that the MMSE function (even less the MMSE matrix) of this sequence of estimators converges as $n \to \infty$. On the other hand, our main interest lies in the mutual information performance of this estimator, in the limit, and the $\limsup$ of the MMSE performance will suffice for this purpose, as shown in Section II-B (see (10) - (12)). As we consider a sequence of optimal estimators, for which the MMSE function can be calculated and is always a non-negative scalar, it defines a total order over such a sequence. Given this order we can consider the $\limsup$ (or $\liminf$) of the sequence of estimators. Note that the supremum is taken with respect to the total order that exists between these estimators - the MMSE function (a scalar quantity), and thus is well-defined. Note however that this definition of the $\limsup$ regards only the performance of the estimators and does not in any sense refer to any ordering between the actual estimators themselves (similar to taking the $\arg\max$ of a set with respect to some performance function). Finally, note that since in (42) we have a limit the $\limsup$ with respect to the total order defined by the MMSE function equals that limit (convergence *a.s.* to some estimator means also the convergence of the MMSE function to the performance of that estimator). We thus take the $\limsup$ of (35):

$$\begin{aligned}
\limsup_{n\to\infty} \mathsf{E}\left\{\boldsymbol{W}_n|\boldsymbol{Y}_n(\gamma)\right\} &= \limsup_{n\to\infty} \mathsf{E}\left\{\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma),\boldsymbol{Z}_n\right\}|\boldsymbol{Y}_n(\gamma)\right\} + \limsup_{n\to\infty} \mathsf{E}\left\{\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\} \\
&\overset{a}{=} \limsup_{n\to\infty} \mathsf{E}\left\{\frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Y}_n(\gamma)-\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n\right)\middle|\boldsymbol{Y}_n(\gamma)\right\} + \limsup_{n\to\infty} \mathsf{E}\left\{\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\} \\
&= \limsup_{n\to\infty} \mathsf{E}\left\{\frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Y}_n(\gamma)-\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n\right)+\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n\middle|\boldsymbol{Y}_n(\gamma)\right\} \\
&= \limsup_{n\to\infty} \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}_n(\gamma) + \limsup_{n\to\infty} \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\} \\
&\overset{b}{=} \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}(\gamma) + \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}|\boldsymbol{Y}(\gamma)\right\}
\end{aligned} \tag{43}$$

where transition a is due to (42) and in transition b we use the following definition:

$$\mathsf{E}\left\{\boldsymbol{Z}|\boldsymbol{Y}(\gamma)\right\} = \limsup_{n\to\infty} \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n\right\} \tag{44}$$

where as in the above the $\limsup$ is defined with respect to the total order defined by the MMSE function.

Note that if the sequence of optimal estimators of $\boldsymbol{W}_n$ converges as $n \to \infty$, it converges to the above estimator.

Now let us consider the following proposed finite estimator:

$$\hat{\boldsymbol{W}}_n = \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}_n(\gamma) + \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}. \tag{45}$$

Taking the $\limsup$ over this sequence of estimators (with respect to the total order defined by the MMSE function of these estimators), we have that

$$\limsup_{n\to\infty} \hat{\boldsymbol{W}}_n = \limsup_{n\to\infty} \mathsf{E}\left\{\boldsymbol{W}_n|\boldsymbol{Y}_n(\gamma)\right\} = \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}(\gamma) + \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}|\boldsymbol{Y}(\gamma)\right\}. \tag{46}$$

Note that we have two sequences of estimators: one is the sequence of optimal estimators for every n, and the other is $\hat{\boldsymbol{W}}_n$, which is sub-optimal for any finite n. Both sequences have the same $\limsup$s convergence meaning

that the $\limsup$ of their MMSE performances are also the same:

$$\limsup_{n\to\infty} \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}_{\boldsymbol{W_n}}(\hat{\boldsymbol{W}}_n)\right) = \limsup_{n\to\infty} \mathsf{MMSE}(\boldsymbol{W}_n|\boldsymbol{Y}_n(\gamma)). \tag{47}$$

We can calculate the mean square error (MSE) matrix of the proposed estimator, denoted by $\mathbf{E}_{\boldsymbol{W_n}}(\hat{\boldsymbol{W}}_n)$ (details are given in Appendix C), and after taking the $\limsup$ of its normalized trace we obtain the following:

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}_{\boldsymbol{W_n}}(\hat{\boldsymbol{W}}_n)\right)\\
&= \frac{\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1} + \frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2}\mathsf{MMSE}(\boldsymbol{Z}|\boldsymbol{Y}(\gamma)) \qquad (48)\\
&+\limsup_{n\to\infty}\frac{1}{n}\mathsf{Tr}\Bigg(\mathsf{E}\Bigg\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma)-\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\Bigg\}\Bigg)\\
&+\limsup_{n\to\infty}\frac{1}{n}\mathsf{Tr}\Bigg(\mathsf{E}\Bigg\{\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma)-\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)^{\mathsf{T}}\Bigg\}\Bigg).\\
&= \frac{\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1} + \frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2}\mathsf{MMSE}(\boldsymbol{Z}|\boldsymbol{Y}(\gamma)) = \mathsf{MMSE}(\boldsymbol{W}|\boldsymbol{Y}(\gamma)) \qquad (49)
\end{aligned}$$

where in the last transition we claim that the last two terms go to zero as $n\to\infty$. This is due to the orthogonality property of the optimal estimator. The details of this step are given in Appendix D. Using (49) in (32) we conclude our proof. ■

Note that Theorem 5 gives us a mutual information - MMSE (I-MMSE) like relationship (see $e.g.$ [55]) for the quantity $I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right)$, although the additive noise is not i.i.d. Gaussian, but rather $\sqrt{\gamma\mathsf{snr}_1}\boldsymbol{X}+\boldsymbol{N}$. We show that precisely in the proof of Theorem 4 given next.

Proof of Theorem 4: When $\gamma \geq 1$ the result is a direct consequence of the assumption that $\{\boldsymbol{X}_n\}_{n\geq 1}$ defines a code sequence of maximum rate, reliably decoded at $\boldsymbol{Y}(1)$ and the chain rule of mutual information (31).

As for $\gamma \in [0,1)$, given Theorem 5 and the chain rule of differentiation, we have that for all independent random processes over $\{\boldsymbol{X}_n,\boldsymbol{Z}_n\}_{n\geq 1}$, both of bounded variances, such that $\{\boldsymbol{X}_n\}_{n\geq 1}$ defines a code sequence of maximum rate $\frac{1}{2}\mathsf{log}(1+\mathsf{snr}_1)$, reliably decoded at $\boldsymbol{Y}(1)$

$$\begin{aligned}
\mathsf{MMSE}\left(\boldsymbol{Z}|\boldsymbol{Y}(\gamma)\right)\cdot\frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2} &= 2\frac{\mathsf{d}}{\mathsf{d}\gamma}I\left(\boldsymbol{Z};\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}+\sqrt{\gamma\mathsf{snr}_1}\boldsymbol{X}+\boldsymbol{N}\right) \qquad (50)\\
&= 2\frac{\mathsf{d}}{\mathsf{d}\gamma'}I\left(\boldsymbol{Z};\sqrt{\gamma'}\boldsymbol{Z}+\sqrt{\frac{\gamma\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}}\boldsymbol{X}+\sqrt{\frac{1}{1+\gamma\mathsf{snr}_1}}\boldsymbol{N}\right)\cdot\frac{\mathsf{d}\gamma'}{\mathsf{d}\gamma}
\end{aligned}$$

where we have defined

$$\gamma' \equiv \frac{\gamma a\mathsf{snr}_2}{1+\gamma\mathsf{snr}_1} \tag{51}$$

and

$$\frac{\mathsf{d}\gamma'}{\mathsf{d}\gamma} = \frac{a\mathsf{snr}_2(1+\gamma\mathsf{snr}_1)-\gamma a\mathsf{snr}_2\mathsf{snr}_1}{(1+\gamma\mathsf{snr}_1)^2} = \frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2}.$$

We further define

$$\boldsymbol{Q}_\gamma \equiv \sqrt{\frac{\gamma\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}}\boldsymbol{X}+\sqrt{\frac{1}{1+\gamma\mathsf{snr}_1}}\boldsymbol{N} \tag{52}$$

which is an additive noise of continuous distribution, since $\boldsymbol{N}$ is normally distributed. Given these definitions the output $\boldsymbol{Y}(\gamma)$ can also be written as

$$\boldsymbol{Y}(\gamma') = \sqrt{\gamma'}\boldsymbol{Z} + \boldsymbol{Q}_{\gamma}. \tag{53}$$

Using these definitions and observations in (50) we have that

$$\frac{\mathsf{d}}{\mathsf{d}\gamma'} I\left(\boldsymbol{Z}; \sqrt{\gamma'}\boldsymbol{Z} + \boldsymbol{Q}_{\gamma}\right) = \frac{1}{2}\mathsf{MMSE}(\boldsymbol{Z}|\sqrt{\gamma'}\boldsymbol{Z} + \boldsymbol{Q}_{\gamma}). \tag{54}$$

Note that if $\boldsymbol{Q}_{\gamma}$ is i.i.d. Gaussian additive noise, for any $\gamma \in [0,1)$ then (54) is the I-MMSE relationship [55]. As $\boldsymbol{N}$ is i.i.d. Gaussian noise, $\boldsymbol{Q}_{\gamma}$ is i.i.d. Gaussian additive noise if and only if $\boldsymbol{X}$ is i.i.d. Gaussian noise. However, this is not the case, as $\boldsymbol{X}_n$ is a codeword from a codebook of rate $\frac{1}{2}\mathsf{log}(1+\mathsf{snr}_1)$. In other words, Theorem 5 results in an I-MMSE-like relationship.

In order to complete the proof of Theorem 4 it remains to show that for all $\gamma \in [0,1)$

$$\mathsf{MMSE}(\boldsymbol{Z}|\sqrt{\gamma'}\boldsymbol{Z} + \boldsymbol{Q}_{\gamma}) = \mathsf{MMSE}(\boldsymbol{Z}|\sqrt{\gamma'}\boldsymbol{Z} + \boldsymbol{N}) \tag{55}$$

where $\boldsymbol{N}$ is an i.i.d. Gaussian random process, meaning that the additive noise $\boldsymbol{Q}_{\gamma}$ has an i.i.d. Gaussian effect on the behavior of the MMSE function (and as such also on the mutual information function), in the limit. Since (54) holds for *independent* random processes over $\{\boldsymbol{X}_n, \boldsymbol{Z}_n\}_{n\geq 1}$ it suffices to examine the behavior of the MMSE assuming a specific random process over $\{\boldsymbol{Z}_n\}_{n\geq 1}$. Note that if $\{\boldsymbol{Z}_n\}_{n\geq 1}$ is an i.i.d. Gaussian process the right-hand side (RHS) of (55) is easily calculated and equals the MSE of the optimal linear estimator

$$\mathsf{MMSE}(\boldsymbol{Z}|\sqrt{\gamma'}\boldsymbol{Z} + \boldsymbol{N}) = \frac{1}{1+\gamma'}. \tag{56}$$

On the other hand, given this assumption on $\{\boldsymbol{Z}_n\}_{n\geq 1}$ we can also calculate the left-hand side (LHS). Recall from (31) that for all $\gamma \in [0,1]$

$$\begin{aligned} I\left(\boldsymbol{W}; \boldsymbol{Y}(\gamma)\right) &= I\left(\boldsymbol{Z}; \boldsymbol{Y}(\gamma)\right) + I\left(\boldsymbol{X}; \boldsymbol{Y}(\gamma)|\boldsymbol{Z}\right) \\ &= I\left(\boldsymbol{Z}; \boldsymbol{Y}(\gamma)\right) + I\left(\boldsymbol{X}; \sqrt{\gamma\mathsf{snr}_1}\boldsymbol{X} + \boldsymbol{N}\right) \\ &= I\left(\boldsymbol{Z}; \boldsymbol{Y}(\gamma)\right) + \frac{1}{2}\mathsf{log}\left(1+\gamma\mathsf{snr}_1\right). \end{aligned} \tag{57}$$

Similarly we also have

$$\begin{aligned} I\left(\boldsymbol{W}; \boldsymbol{Y}(\gamma)\right) &= I\left(\boldsymbol{X}; \boldsymbol{Y}(\gamma)\right) + I\left(\boldsymbol{X}; \boldsymbol{Y}(\gamma)|\boldsymbol{X}\right) \\ &= I\left(\boldsymbol{X}; \boldsymbol{Y}(\gamma)\right) + I\left(\boldsymbol{Z}; \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z} + \boldsymbol{N}\right). \end{aligned} \tag{58}$$

Putting the two together we have that for all $\gamma \in [0,1]$

$$I\left(\boldsymbol{Z}; \boldsymbol{Y}(\gamma)\right) = I\left(\boldsymbol{X}; \boldsymbol{Y}(\gamma)\right) + I\left(\boldsymbol{Z}; \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z} + \boldsymbol{N}\right) - \frac{1}{2}\mathsf{log}\left(1+\gamma\mathsf{snr}_1\right). \tag{59}$$

Since $\boldsymbol{Z}$ is assumed to be i.i.d. Gaussian

$$I\left(\boldsymbol{Z}; \boldsymbol{Y}(\gamma)\right) = I\left(\boldsymbol{X}; \sqrt{\gamma\mathsf{snr}_1}\boldsymbol{X} + \sqrt{1+\gamma a\mathsf{snr}_2}\boldsymbol{N}\right) + \frac{1}{2}\mathsf{log}\left(1+\gamma a\mathsf{snr}_2\right) - \frac{1}{2}\mathsf{log}\left(1+\gamma\mathsf{snr}_1\right). \tag{60}$$

Since $\boldsymbol{X}$ is an optimal point-to-point codebook, its behavior through an AWGN channel is well known and we have that for all $\gamma \in [0,1]$

$$\begin{aligned} I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right) &= \frac{1}{2}\mathsf{log}\left(1+\frac{\gamma\mathsf{snr}_1}{1+\gamma a\mathsf{snr}_2}\right)+\frac{1}{2}\mathsf{log}\left(1+\gamma a\mathsf{snr}_2\right)-\frac{1}{2}\mathsf{log}\left(1+\gamma\mathsf{snr}_1\right) \\ &= \frac{1}{2}\mathsf{log}\left(1+\frac{\gamma a\mathsf{snr}_2}{1+\gamma\mathsf{snr}_1}\right)=\frac{1}{2}\mathsf{log}(1+\gamma'). \end{aligned} \tag{61}$$

Taking the derivative according to γ' in the equivalent region and using (54) we have that

$$\frac{\mathsf{d}}{\mathsf{d}\gamma'}I\left(\boldsymbol{Z};\sqrt{\gamma'}\boldsymbol{Z}+\boldsymbol{Q}_\gamma\right)=\frac{1}{2}\mathsf{MMSE}(\boldsymbol{Z}|\sqrt{\gamma'}\boldsymbol{Z}+\boldsymbol{Q}_\gamma)=\frac{1}{2}\frac{1}{1+\gamma'}. \tag{62}$$

This equals the RHS (56) meaning we have equality in (55) when we assume that $\{\boldsymbol{Z}_n\}_{n\geq 1}$ is i.i.d. Gaussian. Thus, showing that in terms of the MMSE function in the limit as $n\to\infty$ the effect of $\boldsymbol{Q}_\gamma$ is as if it were i.i.d. Gaussian noise. Finally, in order to obtain the expression in the theorem we may apply again the chain rule of derivation as in (50). This concludes the proof. ■

VI. The Costa Conjecture

Let us now consider the two-user Gaussian interference channel:

$$\begin{aligned} \boldsymbol{Y}_{1,n} &= \sqrt{\mathsf{snr}_1}\boldsymbol{X}_n+\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n+\boldsymbol{N}_{1,n} \\ \boldsymbol{Y}_{2,n} &= \sqrt{b\mathsf{snr}_1}\boldsymbol{X}_n+\sqrt{\mathsf{snr}_2}\boldsymbol{Z}_n+\boldsymbol{N}_{2,n}, \end{aligned} \tag{63}$$

where $\boldsymbol{N}_{1,n}$ and $\boldsymbol{N}_{2,n}$ represent standard additive Gaussian noise vectors with independent components. Note that the capacity region of the above model depends on the marginals of the joint conditional distribution of the outputs given the inputs [19], as there is no cooperation between the receivers. As such, we may assume, without loss of generality, $\boldsymbol{N_1}_n=\boldsymbol{N_2}_n=\boldsymbol{N}_n$. $\boldsymbol{X}_n$ and $\boldsymbol{Z}_n$ are assumed to be independent of each other and independent of the AWGN vectors. The average power constraint on both is 1; and $0<a<1$ and $0\leq b$. We assume maximum rate for the transmission of $\boldsymbol{X}_n$, meaning it constitute a "good" code sequence with reliable decoding from the output of an AWGN channel at snr_1. We further assume that $\{\boldsymbol{X}_n,\boldsymbol{Z}_n\}_{n\geq 1}$ have bounded variances. For this setting we can apply Theorem 4 and obtain the next result:

Theorem 6: For any pair of code sequences, $\{\boldsymbol{X}_n,\boldsymbol{Z}_n\}_{n\geq 1}$, both of bounded variances, for which $\{\boldsymbol{X}_n\}_{n\geq 1}$ is a "good" code sequence (attains the maximum possible rate) and reliably decoded from $\boldsymbol{Y}_1$ (63), we have that, in the limit, as $n\to\infty$

$$\mathsf{MMSE}(\boldsymbol{Z}|\sqrt{\mathsf{snr}}\boldsymbol{Z}+\boldsymbol{N})=0,\quad \forall\mathsf{snr}\geq\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}. \tag{64}$$

Note that as in the previous section, if the above MMSE function does not converge as $n\to\infty$, the above theorem applies to the $\limsup$ of the MMSE sequence. Furthermore, note that we do not require in the above decoding requirements on the transmission in $\{\boldsymbol{Z}_n\}_{n\geq 1}$, but only the reliable decoding of $\{\boldsymbol{X}_n\}_{n\geq 1}$.

Proof: Considering the limiting expression for the capacity of the two-user Gaussian interference channel [2], we have

$$\cup_{P_{\boldsymbol{Z}},P_{\boldsymbol{X}}}\begin{cases}\mathsf{R}_x & \leq I\left(\boldsymbol{X};\boldsymbol{Y}_1\right)=I\left(\boldsymbol{X};\boldsymbol{Y}(1)\right)\\ \mathsf{R}_z & \leq I\left(\boldsymbol{Z};\boldsymbol{Y}_2\right)\end{cases}.$$

Let us assume that we have a pair of code sequences for the Gaussian interference channel. We assume that at $\boldsymbol{Y}(1)$ we can decode $\boldsymbol{X}$ reliably, meaning,

$$\begin{aligned}I\left(\boldsymbol{Z};\boldsymbol{Y}(1)\right)&=I\left(\boldsymbol{Z};\boldsymbol{Y}(1),\boldsymbol{X}\right)\\&=I\left(\boldsymbol{Z};\boldsymbol{Y}(1)|\boldsymbol{X}\right)\\&=I\left(\boldsymbol{Z};\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}+\boldsymbol{N}\right).\end{aligned}$$

On the other hand, we may apply Theorem 4 and conclude that

$$\begin{aligned}I\left(\boldsymbol{Z};\boldsymbol{Y}(1)\right)&=\frac{1}{2}\int_0^1\mathsf{MMSE}(\boldsymbol{Z}|\sqrt{\frac{\gamma a\mathsf{snr}_2}{1+\gamma\mathsf{snr}_1}}\boldsymbol{Z}+\boldsymbol{N})\cdot\frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2}\mathsf{d}\gamma\\&=I\left(\boldsymbol{Z};\sqrt{\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}}\boldsymbol{Z}+\boldsymbol{N}\right)\end{aligned}\tag{65}$$

meaning that, for any such pair of reliable code sequences, we have the following equality:

$$\begin{aligned}I\left(\boldsymbol{Z};\boldsymbol{Y}(1)\right)&=I\left(\boldsymbol{Z};\sqrt{a\mathsf{snr}_2}\boldsymbol{Z}+\boldsymbol{N}\right)\\&=I\left(\boldsymbol{Z};\sqrt{\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}}\boldsymbol{Z}+\boldsymbol{N}\right)\end{aligned}\tag{66}$$

which by the I-MMSE relationship [55] means that the MMSE when estimating $\boldsymbol{Z}$ from the output of an AWGN channel is zero for any such code sequences for all SNRs in the region $\left(\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1},a\mathsf{snr}_2\right)$, and hence any SNR $\geq\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}$. This concludes our proof. ■

The following corollary follows immediately:

Corollary 1: Under the assumption of Theorem 6 we have that

$$2\frac{\mathsf{d}}{\mathsf{d}\gamma}I\left(\boldsymbol{Z};\boldsymbol{Y}(\gamma)\right)=\begin{cases}\mathsf{MMSE}\left(\boldsymbol{Z}|\sqrt{\frac{\gamma a\mathsf{snr}_2}{1+\gamma\mathsf{snr}_1}}\boldsymbol{Z}+\boldsymbol{N}\right)\cdot\frac{a\mathsf{snr}_2}{(1+\gamma\mathsf{snr}_1)^2}, & \gamma\in[0,1)\\ 0, & \gamma\geq1\end{cases}.\tag{67}$$

The above result provides us with corner points of the capacity region of the two-user Gaussian interference channel, under the conditions stated in Theorem 6. Depending on the value of b we have three different types of interference channels: $b=0$ is a one-sided (or Z) interference channel, $b\in(0,1)$ is the weak interference channel and $b\geq1$ is the mixed interference channel. Note that for the reverse case, in which $\boldsymbol{Z}$ is transmitted at maximum rate, the corner points of the capacity region of both the one-sided case and the mixed case are known, and are the corners yielding sum-capacity [16], [30], [32].

Corollary 2: For the one-sided Gaussian interference channel ((63) with $b=0$), when the input distributions are of bounded variances, we have the following corner point:

$$\left(\frac{1}{2}\log\left(1+\mathsf{snr}_1\right),\frac{1}{2}\log\left(1+\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}\right)\right).\tag{68}$$

Proof: Using Theorem 6 the solution to the maximization yields, therefore,

$$\max I\left(\boldsymbol{Z};\sqrt{\mathsf{snr}_2}\boldsymbol{Z}+\boldsymbol{N}\right)=\frac{1}{2}\log\left(1+\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}\right), \tag{69}$$

thus, concluding our proof. ■

Note, that the above optimization problem can also be written as follows:

$$\begin{aligned} \max \quad & I\left(\boldsymbol{Z};\sqrt{\mathsf{snr}_2}\boldsymbol{Z}+\boldsymbol{N}\right) \\ \text{s.t.} \quad & \mathsf{MMSE}(\boldsymbol{Z}|\boldsymbol{Y}(1))=0 \end{aligned} \tag{70}$$

for which the solution is according to the I-MMSE trade-off shown in [40], with the assumption that the additive noise $\boldsymbol{N}+\sqrt{\mathsf{snr}_1}\boldsymbol{X}$ is i.i.d. Gaussian. Moreover, we do not need to limit this optimization problem to an average power constraint on $\boldsymbol{Z}$ and can associate it with other constraints on $\boldsymbol{Z}$, such as a peak amplitude constraint.

Theorem 7: For the weak two-user Gaussian interference channel ((63) with $b\in(0,1)$), when the input distributions are of bounded variances, we have the following corner points:

$$\left(\frac{1}{2}\log\left(1+\mathsf{snr}_1\right),\frac{1}{2}\log\left(1+\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}\right)\right) \text{ and} \tag{71}$$

$$\left(\frac{1}{2}\log\left(1+\frac{b\mathsf{snr}_1}{1+\mathsf{snr}_2}\right),\frac{1}{2}\log\left(1+\mathsf{snr}_2\right)\right). \tag{72}$$

Proof: Note that we may apply Theorem 4 to conclude that

$$2\frac{\mathsf{d}}{\mathsf{d}\gamma}I\left(\boldsymbol{Z};\sqrt{\gamma b\mathsf{snr}_1}\boldsymbol{X}+\sqrt{\gamma\mathsf{snr}_2}\boldsymbol{Z}+\boldsymbol{N}_2\right)=\mathsf{MMSE}\left(\boldsymbol{Z}|\sqrt{\frac{\gamma\mathsf{snr}_2}{1+\gamma b\mathsf{snr}_1}}\boldsymbol{Z}+\boldsymbol{N}\right)\cdot\frac{\mathsf{snr}_2}{(1+\gamma b\mathsf{snr}_1)^2} \tag{73}$$

for all $\gamma\in[0,1]$ (simply denote redefine snr_1 n the Theorem as $b\mathsf{snr}_1$ and $a\mathsf{snr}_2$ as snr_2). Integrating over the above result we have

$$I\left(\boldsymbol{Z};\boldsymbol{Y}_2\right)=I\left(\boldsymbol{Z};\sqrt{\frac{\mathsf{snr}_2}{1+b\mathsf{snr}_1}}\boldsymbol{Z}+\boldsymbol{N}\right). \tag{74}$$

Thus, maximizing $I\left(\boldsymbol{Z};\boldsymbol{Y}_2\right)$ results in

$$\min\left\{\frac{1}{2}\log\left(1+\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}\right),\frac{1}{2}\log\left(1+\frac{\mathsf{snr}_2}{1+b\mathsf{snr}_1}\right)\right\}=\frac{1}{2}\log\left(1+\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}\right) \tag{75}$$

due to Theorem 6 and the maximum entropy result. This provides us with (71). (72) is simply the symmetric result and can be derived in a same manner. This concludes our proof. ■

Finally, we have the following result for the mixed case.

Theorem 8: For the mixed two-user Gaussian interference channel ((63) with $b\geq 1$), when the input distributions are of bounded variances, we have the following corner point:

$$\left(\frac{1}{2}\log\left(1+\mathsf{snr}_1\right),\frac{1}{2}\log\left(1+\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}\right)\right). \tag{76}$$

Proof: Since $b\geq 1$ and $\boldsymbol{Z}$ can be reliably decoded at $\boldsymbol{Y}_2$, we can also reliably decode $\boldsymbol{X}$ at $\boldsymbol{Y}_2$. Thus, a trivial outer bound on this corner point is the MAC corner point:

$$\left(\frac{1}{2}\log\left(1+\mathsf{snr}_1\right),\frac{1}{2}\log\left(\frac{1+\mathsf{snr}_2+b\mathsf{snr}_1}{1+\mathsf{snr}_1}\right)\right). \tag{77}$$

However, we also have that

$$I\left(\boldsymbol{Z};\boldsymbol{Y}_2\right) \leq I\left(\boldsymbol{Z};\sqrt{\mathsf{snr}_2}\boldsymbol{Z}+\boldsymbol{N}\right) \tag{78}$$

and the maximization of the RHS of this inequality using Theorem 6 gives us

$$I\left(\boldsymbol{Z};\boldsymbol{Y}_2\right) \leq I\left(\boldsymbol{Z};\sqrt{\mathsf{snr}_2}\boldsymbol{Z}+\boldsymbol{N}\right) \leq \frac{1}{2}\mathsf{log}\left(1+\frac{a\mathsf{snr}_2}{1+\mathsf{snr}_1}\right).$$

This upper bound is tighter than the MAC outer bound. It can be shown that this corner point can be achieved using Gaussian point-to-point codes and joint decoding, as shown in [45]. This concludes our proof. ■

Thus, treating the interference as Gaussian noise when $b \in [1, \frac{1-a+\mathsf{snr}_1}{a\mathsf{snr}_1}]$ is optimal, not only in terms of the GDoF, as recently shown in [74]. Moreover, note that non-unique decoding at $\boldsymbol{Y}_2$ [51], which potentially could obtain higher achievable rates, does not improve the above corner point. Note further that for any pair of code sequences (of bounded variances)that achieves these corner points, both messages can be reliably decoded. Moreover, at $\boldsymbol{Y}_1$ the decoding of both messages can always be done sequentially, due to Theorem 6, which guarantees MMSE of zero for the estimation of $\boldsymbol{Z}$ while considering $\boldsymbol{X}$ as additive Gaussian noise.

VII. Discussion and Conclusions

In this work we have examined the effect of "good" code sequences on an additional transmitted signal, over the additive Gaussian channel. We have shown that a maximum rate transmission creates, effectively, additional additive Gaussian noise, in terms of the mutual information and MMSE. This observation leads to the fact that, if one requires to reliably decode the maximum rate transmission, the MMSE of any additional transmitted signal must be zero. This condition is orthogonal to any additional condition on the rate of this additional transmission. Using these observations we obtain corner points of the two-user Gaussian interference channel. These results hold assuming that the independent inputs are of bounded variances.

It has come to our attention that in parallel to our work Polyanskiy and Wu [75] solved "Costa's conjecture" regarding the corner point of the two-user Gaussian interference cannel, under some regularity conditions. Their solution differs substantially from our own, as they target the original claim in [15] (see also [30]) and show the continuity of the differential entropy.

Appendix

A. Proof of Theorem 2

Proof: We propose an alternative proof from an estimation point of view through the I-MMSE relationship. Proving the theorem is equivalent to showing that in order to obtain the maximum rate, the integrand must equal its maximum value for all $\gamma \in [0, \mathsf{snr}]$. That leads to an exact expression for $\lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n;\sqrt{\gamma}\boldsymbol{X}_n+\boldsymbol{N}_n\right)$ for any γ which is the result in [56], where binary inputs were considered.

We assume an optimal point-to-point code of rate $\frac{1}{2}\mathsf{log}(1+\mathsf{snr})$, meaning that

$$\lim_{n\to\infty}\frac{1}{n}I\left(\boldsymbol{X}_n;\sqrt{\mathsf{snr}}\boldsymbol{X}_n+\boldsymbol{N}_n\right)=\frac{1}{2}\mathsf{log}(1+\mathsf{snr}). \tag{79}$$

Let us first consider the following optimization problem:

$$\begin{aligned}\max \quad & \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}_{\boldsymbol{X_n}}(\gamma)\right) = \mathsf{MMSE}^{\mathsf{c}_n}(\gamma) \\ \text{s.t. } & \frac{1}{n}\mathsf{Tr}(\mathbf{R}_{\boldsymbol{X}}) \leq 1. \end{aligned} \tag{80}$$

This problem was solved in [76, Lemma 2], using the concavity of the function $\frac{\lambda}{1+\gamma\lambda}$ and majorization theory and the solution is a Gaussian distribution over $\boldsymbol{X}_n$ with covariance matrix $\mathbf{R}_{\boldsymbol{X}} = \mathbf{I}$ (uniform distribution of the power) for any n. From this we can conclude that for any $\gamma \geq 0$ under the above trace constraint

$$\mathsf{MMSE}^{\mathsf{c}}(\gamma)_{sup} \leq \frac{1}{1+\gamma}. \tag{81}$$

Thus, given the above and (11) we have that under the trace constraint

$$\begin{aligned}\lim_{n\to\infty}\int_0^\rho \mathsf{MMSE}^{\mathsf{c}_n}(\gamma)\mathsf{d}\gamma &\leq \lim_{n\to\infty}\int_0^\rho \frac{1}{1+\gamma}\mathsf{d}\gamma \\ \int_0^\rho \mathsf{MMSE}^{\mathsf{c}}(\gamma)_{sup}\mathsf{d}\gamma &\leq \int_0^\rho \frac{1}{1+\gamma}\mathsf{d}\gamma\end{aligned}$$

with equality if and only if

$$\mathsf{MMSE}^{\mathsf{c}}(\gamma)_{sup} = \frac{1}{1+\gamma}, \quad \forall\gamma \in [0,\rho). \tag{82}$$

As such, obtaining capacity at snr requires the above equality for all $\gamma \in [0,\mathsf{snr})$ and thus,

$$\frac{1}{2}\int_0^\rho \mathsf{MMSE}^{\mathsf{c}}(\gamma)_{sup}\mathsf{d}\gamma = \frac{1}{2}\mathsf{log}(1+\rho), \quad \forall\rho \in [0,\mathsf{snr}). \tag{83}$$

Finally, note that the theorem regards any optimal point-to-point codebook, that attains capacity and complies with the power constraint, that is, any codeword $\boldsymbol{x}_n \in \mathsf{C}_n$ has limited power:

$$\frac{1}{n}||\boldsymbol{x}_n||^2 \leq 1. \tag{84}$$

The normalized trace of the covariance matrix of such codes can be written as follows:

$$\frac{1}{n}\mathsf{Tr}\left(\mathbf{R}_{\boldsymbol{X_n}}\right) = \mathsf{E}\left\{\frac{1}{n} \parallel \boldsymbol{X}_n \parallel^2\right\} \leq 1 \tag{85}$$

where the inequality is a direct consequence of the power constraint (84). Thus, the power constraint on the code leads to the following constraint:

$$\frac{1}{n}\mathsf{Tr}(\mathbf{R}_{\boldsymbol{X_n}}) \leq 1, \quad \forall n \geq 1. \tag{86}$$

This concludes our proof. ■

B. Proof of Theorem 3

Proof: We first consider the MSE matrix of the optimal bit-wise linear estimator, which we denote by $\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)$. Since each codeword complies with the power constraint (84) we have that

$$\begin{aligned}\frac{1}{n}\mathsf{Tr}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) &= \frac{1}{n}\sum_n \frac{M_i}{1+\gamma\mathsf{snr}M_i} \\ &\leq \frac{1}{1+\gamma\mathsf{snr}}\end{aligned} \tag{87}$$

where $M_i = \mathsf{E}\left\{X_i^2\right\}$ (the power of the i^{th} index), and the inequality is due to the strict concavity of the function $\frac{\lambda}{1+\gamma\lambda}$, Jensen's inequality and the fact that it is monotonically increasing. Note that the RHS does not depend on n, and thus in taking the limit (or alternatively, the $\limsup$) the RHS is unaffected.

For any n the MMSE matrix can be upper bounded (in the positive semidefinite sense) by the MSE of the optimal bit-wise linear estimator:

$$\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) \preceq \mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n), \quad \forall n \geq 0 \tag{88}$$

and

$$\frac{1}{n}\mathsf{Tr}\left(\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) \preceq \frac{1}{n}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right), \quad \forall n \geq 0. \tag{89}$$

Recall our assumption in (17):

$$\limsup_{n\rightarrow\infty} \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) = \frac{1}{1+\gamma\mathsf{snr}}. \tag{90}$$

Using (87) and (89) we can conclude that

$$\begin{aligned}\limsup_{n\rightarrow\infty} \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) &= \frac{1}{1+\gamma\mathsf{snr}} \\ &\leq \limsup_{n\rightarrow\infty} \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) \\ &\leq \frac{1}{1+\gamma\mathsf{snr}},\end{aligned} \tag{91}$$

meaning we have equality in the above

$$\limsup_{n\rightarrow\infty} \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) = \frac{1}{1+\gamma\mathsf{snr}}. \tag{92}$$

As shown also in [58] the optimization problem

$$\begin{aligned}\max \quad & \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) \\ \text{s.t.} \quad & \frac{1}{n}\mathsf{Tr}\left(\mathbf{R}_{\boldsymbol{X_n}}\right) \leq 1\end{aligned} \tag{93}$$

is strictly concave, with a solution for every n being $\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) = \frac{1}{1+\gamma\mathsf{snr}}\mathbf{I}_n$. We define total order between every two matrices according to the following distance measure

$$d\left(\mathbf{A}_n, \mathbf{B}_n\right) = \frac{1}{n}\mathsf{Tr}\left(\mathbf{A}_n - \mathbf{B}_n\right). \tag{94}$$

Due to the strict concavity of the optimization problem and the equality in (92), suggesting that in the limit the matrices are within some sleeve bounded by the matrix $\frac{1}{1+\gamma\mathsf{snr}}\mathbf{I}_n$ obtaining the optimum solution (a sleeve defined using our distance measure). Thus apply the $\limsup$ with respect to the defined total order and attain the following

$$\limsup_{n\rightarrow\infty} \mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) = \frac{1}{1+\gamma\mathsf{snr}}\mathbf{I}. \tag{95}$$

Note that this does not mean that $\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)$ converges to $\frac{1}{1+\gamma\mathsf{snr}}\mathbf{I}$ as we are considering only the $\limsup$. Moreover, the claim states that the matrix is within a sleeve bounded by the matrix $\frac{1}{1+\gamma\mathsf{snr}}\mathbf{I}$ in terms of the distance function d (94). This can also be written as follows: for any $\epsilon > 0$ there exists an N_ϵ such that

$$\max_{m\geq n} d\left(\mathbf{E}^{OBL}_{\boldsymbol{X_M}}(\boldsymbol{Y}^c_m), \frac{1}{1+\gamma\mathsf{snr}}\mathbf{I}_m\right) \leq \epsilon, \quad \forall n \geq N_\epsilon. \tag{96}$$

Note that this a consequence of the strict concavity of the problem and limited to covariance matrices that comply with the trace constraint. Note also that for a sequence of positive semi-definite matrices the $\limsup$ and the $\liminf$, defined according to the total order defined by the function d, results with a positive semi-definite matrix, thus due to (88) we have that

$$\begin{aligned}\limsup_{n\to\infty}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_M}}(\boldsymbol{Y}^c_m) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) &\succeq \mathbf{0}\\ \liminf_{n\to\infty}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_M}}(\boldsymbol{Y}^c_m) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) &\succeq \mathbf{0}.\end{aligned} \tag{97}$$

In other words for any sequence of vectors $\mathbf{a}_n \in \mathbb{R}^n$

$$\begin{aligned}\limsup_{n\to\infty}\mathbf{a}_n^\mathsf{T}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right)\mathbf{a}_n &\geq 0\\ \liminf_{n\to\infty}\mathbf{a}_n^\mathsf{T}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right)\mathbf{a}_n &\geq 0.\end{aligned} \tag{98}$$

Moreover (97) also implies that

$$\begin{aligned}\limsup_{n\to\infty}\mathbf{E}^{OBL}_{\boldsymbol{X_M}}(\boldsymbol{Y}^c_m) &\succeq \limsup_{n\to\infty}\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\\ \frac{1}{1+\gamma\mathsf{snr}}\mathbf{I} &\succeq \limsup_{n\to\infty}\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n).\end{aligned} \tag{99}$$

From this and our assumption (90), meaning that the matrix $\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)$ also attains the maximum value (in terms of d) we can conclude that

$$\limsup_{n\to\infty}\left[\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right]_{ii} = \frac{1}{1+\gamma\mathsf{snr}} \tag{100}$$

where the $\limsup$ is according to the total order defined by the function d (although we are considering a scalar quantity here). Alternatively, if this does not hold (90) could not be attained given the requirement in (99). Given this observation and (98) we have that for any sequence of vectors $\mathbf{a}_n \in \mathbb{R}^n$

$$\limsup_{n\to\infty}\mathbf{a}_n^\mathsf{T}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right)\mathbf{a}_n = 0. \tag{101}$$

This is equivalent to

$$\limsup_{n\to\infty}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) = \mathbf{0}. \tag{102}$$

Since

$$\limsup_{n\to\infty}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) \not\succeq \liminf_{n\to\infty}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n) - \mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right) \tag{103}$$

and due to (97) we have also that

$$\liminf_{n\to\infty}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)-\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right)=\mathbf{0}. \tag{104}$$

Thus the limit of this difference exists and is

$$\lim_{n\to\infty}\left(\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)-\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right)=\mathbf{0}. \tag{105}$$

Note that we have show only the existence of the limit of the difference between these two matrices and do not claim anything about the convergence of either one.

We now turn to consider the sequence of optimal estimators. The optimal estimator is *a.s.* unique for any n including $n\to\infty$. This means that the sequence of optimal estimators converges *a.s.*. Consider the difference in performance (per-component) between the optimal estimator and the optimal bit-wise linear estimator

$$\begin{aligned}
&\left|\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-X\right)^2\right\}-\mathsf{E}\left\{\left(\hat{X}^{OBL}(Y^c)-X\right)^2\right\}\right|\\
&=\left|\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)+\hat{X}^{OBL}(Y^c)-X\right)^2\right\}-\mathsf{E}\left\{\left(\hat{X}^{OBL}(Y^c)-X\right)^2\right\}\right|\\
&=\left|\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)\right)^2\right\}+2\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)\right)\left(\hat{X}^{OBL}(Y^c)-X\right)\right\}\right|\\
&=\left|\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)\right)^2\right\}\right.\\
&\qquad\left.+2\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)\right)\left(\hat{X}^{OBL}(Y^c)-\hat{X}(\boldsymbol{Y}^c_n)+\hat{X}(\boldsymbol{Y}^c_n)-X\right)\right\}\right|\\
&=\left|-\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)\right)^2\right\}+2\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)\right)\left(\hat{X}(\boldsymbol{Y}^c_n)-X\right)\right\}\right|\\
&\overset{a}{=}\left|-\mathsf{E}\left\{\left(\hat{X}(\boldsymbol{Y}^c_n)-\hat{X}^{OBL}(Y^c)\right)^2\right\}\right|
\end{aligned} \tag{106}$$

where transition a is due to the orthogonality of the estimation error with every function of $\boldsymbol{Y}_n$. From the convergence of the difference between the two MSE matrices we can conclude that a mean square convergence of the optimal estimator to the optimal bit-wise linear estimator. Thus, we can conclude that the *a.s.* convergence of the optimal estimator is to the optimal bit-wise linear estimator.

Note that as we have concluded (100) we can also conclude the following

$$\limsup_{n\to\infty}\left[\mathbf{E}^{OBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}^c_n)\right]_{ii}=\frac{1}{1+\gamma\mathsf{snr}} \tag{107}$$

where again the $\limsup$ is according to the total order defined by d. The meaning of this is that if we examine the sequence of optimal bit-wise estimators per-component as $n\to\infty$ we observe that they are also limited to some sleeve around the following bit-wise estimator

$$\frac{\sqrt{\gamma\mathsf{snr}}}{1+\gamma\mathsf{snr}}Y^c \tag{108}$$

in terms of their performance as defined by the function d (94). Thus, we can further conclude that the *a.s.* convergence of the above two estimators (the optimal estimator and the optimal bit-wise estimator) is to (108).

This conclude the proof. ■

C. Derivation of Equation 48

We can calculate the MSE matrix of the proposed estimator, that is,

$$\mathbf{E}_{\boldsymbol{W}_n}(\hat{\boldsymbol{W}}_n) = \mathsf{E}\left\{\left(\boldsymbol{W}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}_n(\gamma) - \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\left(\boldsymbol{W}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}_n(\gamma) - \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)^{\mathsf{T}}\right\} \tag{109}$$

We first observe that

$$\begin{aligned}
&\boldsymbol{W}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}_n(\gamma) - \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\\
&= \sqrt{\mathsf{snr}_1}\boldsymbol{X}_n + \sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\boldsymbol{Y}_n(\gamma) - \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\\
&= \sqrt{\mathsf{snr}_1}\boldsymbol{X}_n + \sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n) - \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\\
&= \sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n) + \sqrt{a\mathsf{snr}_2}\boldsymbol{Z}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}\sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n - \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\\
&= \sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n) + \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\boldsymbol{Z}_n - \frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\} \qquad (110)
\end{aligned}$$

and thus,

$$\begin{aligned}
&\mathbf{E}_{\boldsymbol{W}_n}(\hat{\boldsymbol{W}}_n)\\
&= \mathsf{E}\left\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_{\boldsymbol{W},\boldsymbol{n}}(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)^{\mathsf{T}}\right\}\\
&\quad + \mathsf{E}\left\{\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&\quad + \mathsf{E}\left\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&\quad + \mathsf{E}\left\{\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)^{\mathsf{T}}\right\}\\
&= \mathsf{E}\left\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \sqrt{\mathsf{snr}_1}\hat{\boldsymbol{X}}_n^{SBL}\right)\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \sqrt{\mathsf{snr}_1}\hat{\boldsymbol{X}}_n^{SBL}\right)^{\mathsf{T}}\right\}\\
&\quad + \mathsf{E}\left\{\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&\quad + \mathsf{E}\left\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&\quad + \mathsf{E}\left\{\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)^{\mathsf{T}}\right\}. \qquad (111)
\end{aligned}$$

In order to conclude we need the following result regarding the performance of the sub-optimal bit-wise linear estimator:

Corollary 3: Considering $\hat{X}^{SBL}$ we have the following properties for its performance:

$$\limsup_{n\to\infty} \left[\mathbf{E}^{SBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n, \boldsymbol{Z}_n)\right]_{ii} = \frac{1}{1+\gamma\mathsf{snr}_1}, \text{ and}$$
$$\limsup_{n\to\infty} \frac{1}{n}\mathsf{Tr}\left(\mathbf{E}^{SBL}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n, \boldsymbol{Z}_n)\right) = \frac{1}{1+\gamma\mathsf{snr}_1}. \tag{112}$$

Proof: Both claims are a direct result of Theorem 3 where we have shown that the optimal bit-wise estimator converges to the sub-optimal bit-wise estimator and thus its performance in the limit follow. ∎

Thus, taking the $\limsup$ of the normalized trace of the above result we obtain (48).

D. Proof That Cross Terms Tend to Zero

The proof is identical for both terms so we will show it only for the first one.

$$\begin{aligned}
&\mathsf{E}\left\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&= \mathsf{E}\left\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} + \mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\cdot\right.\\
&\left.\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&= \mathsf{E}\left\{\left(\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n - \mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\}\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&+ \mathsf{E}\left\{\left(\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\\
&= \mathsf{E}\left\{\left(\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}
\end{aligned} \tag{113}$$

where the last transition is due to the orthogonality property (the estimation error of the optimal estimator of $\boldsymbol{X}_n$ from $\boldsymbol{Y}_n$ and $\boldsymbol{Z}_n$ is orthogonal to any function of the measurements, for every n). Recall that our goal is to prove that the $\limsup$ of the normalized trace of the above matrix equals zero, that is,

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\mathsf{Tr}\left(\mathsf{E}\left\{\left(\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}\boldsymbol{X}_n|\boldsymbol{Y}_n(\gamma), \boldsymbol{Z}_n\right\} - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(\boldsymbol{Y}_n(\gamma) - \sqrt{\gamma a\mathsf{snr}_2}\boldsymbol{Z}_n)\right)\cdot\right.\right.\\
&\left.\left.\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(\boldsymbol{Z}_n - \mathsf{E}\left\{\boldsymbol{Z}_n|\boldsymbol{Y}_n(\gamma)\right\}\right)\right)^{\mathsf{T}}\right\}\right) = 0.
\end{aligned} \tag{114}$$

Due to the linearity of both the trace function and the expectation, it suffices to show that

$$\limsup_{n\to\infty}\mathsf{E}\left\{\left(\mathsf{E}\left\{\sqrt{\mathsf{snr}_1}X|\boldsymbol{Y}_n, \boldsymbol{Z}_n\right\} - \frac{\sqrt{\gamma}\mathsf{snr}_1}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a\mathsf{snr}_2}Z)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(Z - \mathsf{E}\left\{Z|\boldsymbol{Y}_n\right\}\right)\right)\right\} = 0 \tag{115}$$

which can also be written as

$$\limsup_{n\to\infty}\mathsf{E}\left\{\left(G_n - \frac{\sqrt{\gamma\mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a\mathsf{snr}_2}Z)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}\left(Z - H_n\right)\right)\right\} = 0. \tag{116}$$

Using the Cauchy-Schwarz inequality we have that

$$\left(\mathsf{E}\left\{\left(G_n - \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a \mathsf{snr}_2} Z)\right)\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}(Z-H_n)\right)\right\}\right)^2 \leq \mathsf{E}\left\{\left(G_n - \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a \mathsf{snr}_2} Z)\right)^2\right\}\mathsf{E}\left\{\left(\frac{\sqrt{a\mathsf{snr}_2}}{1+\gamma\mathsf{snr}_1}(Z-H_n)\right)^2\right\}. \quad (117)$$

Note that the second term on the RHS is bounded as it is a factor of the MMSE function when estimating Z from the output $\boldsymbol{Y}_n$ (bounded by the variance on Z which is assumes bounded). It remains to show that the $\limsup$ of the first term is zero, that is,

$$\limsup_{n\to\infty} \mathsf{E}\left\{\left(G_n - \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a \mathsf{snr}_2} Z)\right)^2\right\} = 0, \quad (118)$$

meaning an $\mathcal{L}^2$ convergence of the sequence G_n. Before taking the $\limsup$ we write

$$\begin{aligned}
&\mathsf{E}\left\{\left(G_n - \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a \mathsf{snr}_2} Z)\right)^2\right\} \\
&= \mathsf{E}\left\{\left(G_n - X + X - \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a \mathsf{snr}_2} Z)\right)^2\right\} \\
&= \mathsf{E}\left\{(G_n - X)^2\right\} + \mathsf{E}\left\{\left(X - \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a \mathsf{snr}_2} Z)\right)^2\right\} + 2\mathsf{E}\left\{(G_n - X)\left(X - \frac{\sqrt{\gamma \mathsf{snr}_1}}{1+\gamma\mathsf{snr}_1}(Y - \sqrt{\gamma a \mathsf{snr}_2} Z)\right)\right\} \\
&\stackrel{a}{=} [\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n, \boldsymbol{Z}_n)]_{ii} + \mathsf{E}\left\{\left(X - \hat{X}^{SBL}\right)^2\right\} + 2\mathsf{E}\left\{(G_n - X)\,X\right\} \\
&\stackrel{b}{=} [\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n, \boldsymbol{Z}_n)]_{ii} + \mathsf{E}\left\{\left(X - \hat{X}^{SBL}\right)^2\right\} - 2\mathsf{E}\left\{(G_n - X)\,(G_n - X)\right\} \\
&= [\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n, \boldsymbol{Z}_n)]_{ii} + \mathsf{E}\left\{\left(X - \hat{X}^{SBL}\right)^2\right\} - 2\,[\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n, \boldsymbol{Z}_n)]_{ii} \\
&= \mathsf{E}\left\{\left(X - \hat{X}^{SBL}\right)^2\right\} - [\mathbf{E}_{\boldsymbol{X_n}}(\boldsymbol{Y}_n, \boldsymbol{Z}_n)]_{ii} \qquad (119)
\end{aligned}$$

where i denotes the time index of the component X being estimated. Both transitions a and b are due to the orthogonality property of the optimal MMSE estimator. Taking the $\limsup$ of the above yields zero due to Corollary 3 which shows that the sub-optimal linear estimator has the same performance in this sense as the optimal bit-wise linear estimator and the optimal estimator (see proof of Theorem 3) for a "good" code sequence. This proves that (118) holds.

REFERENCES

[1] C. E. Shannon, "Two-way communication channels," in *Proc. 4^{th} Berkeley Symp. Mathematical Statistics and Probability*, pp. 611–644, Berkeley, CA 1961.

[2] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE International Symposium on Information Theory (ISIT 1971)*, pp. 23–52, Tsahkadsor, Armenia, U.S.S.R, Sptember 1971.

[3] A. B. Carleial, "A case where interference does not reduce capacity," *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 569–570, September 1975.

[4] H. Sato, "The capacity of the Gaussian interference channel under strong interference," *IEEE Transactions on Information Theory*, vol. 27, no. 6, pp. 786–788, November 1981.

[5] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 49–60, January 1981.

[6] M. H. M. Costa and A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference," *IEEE Transactions on Information Theory*, vol. 33, no. 5, pp. 710–711, September 1985.

[7] H. F. Chong, M. Motani, H. K. Grag, and H. El Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 3188–3195, July 2008.

[8] D. Tuninetti, "K-user interference channels: General outer bounds and sum-capacity for certain Gaussian channels," in *Proc. IEEE International Symposium on Information Theory (ISIT 2011)*, Saint Petersburg, Russia, July 31 - August 5 2011.

[9] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5534–5562, December 2008.

[10] C. Nair, L. Xia, and M. Yazdanpanah, "Sub-optimality of the Han-Kobayashi achievable region for interference channels," 2015, arXiv:1502.02589.

[11] X. Shang and B. Chen, "Two-user gaussian interference channels: An information theoretic point of view," *Foundations and Trends in Signal Processing*, vol. 10, no. 3, pp. 247–378, 2013.

[12] A. B. Carleial, "Outer bounds on the capacity of interference channels," *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 602–606, July 1983.

[13] H. Sato, "Two-user communication channels," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 295–304, May 1977.

[14] ——, "On degraded Gaussian two-user channels," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp. 637–640, September 1977.

[15] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Transactions on Information Theory*, vol. 31, no. 5, pp. 607–615, September 1985.

[16] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 620–643, February 2009.

[17] T. Liu and P. Viswanath, "An external inequality motivated by multiterminal information theoretic problems," in *Proc. IEEE International Symposium on Information Theory (ISIT 2008)*, pp. 1016–1020, Seattle, WA, July 9-14 2006.

[18] V. S. Annapureddy and V. Veeravalli, "Gaussian interference networks: Sum capacity in the low interference regime and new upper bounds on the capacity region," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3032–3050, July 2009.

[19] T. M. Cover and J. A. Thomas, *Elements in Information Theory*. New York: Wiley, 2006, second Edition.

[20] R. Bustin and S. Shamai (Shitz), "The I-MMSE approach on the weak Gaussian Z-interference channel and the type I Gaussian broadcast-Z-interference channel," in *Proc. of 2010 IEEE 26-th convention of Electrical and Electronics Engineers in Israel, Eilat, 2010*, November 2010.

[21] O. Rioul and M. H. M. Costa, "Almost there: Corner points of Gaussian interference channels," *Information Theory and Applications Workshop (ITA 2015)*, UCSD, San Diego, CA, USA, February 1 - 6, 2015.

[22] G. Kramer, "Outer bounds on the capacity of Gaussian interference channels," *IEEE Transactions on Information Theory*, vol. 50, no. 3, pp. 581–586, March 2004.

[23] E. Telatar and D. Tse, "Bounds on the capacity region of a class of interference channels," in *Proc. IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, France, June 24 - 29 2007.

[24] C. S. Vaze and M. K. Varanasi, "The capacity region of the symmetric Gaussian interference channel with common information within a constant gap," in *Proc. IEEE International Symposium on Information Theory (ISIT 2011)*, Saint Petersburg, Russia, July 31 - August 5 2011.

[25] S. Karmakar and M. K. Varanasi, "Capacity of the MIMO interference channel to within a constant gap," in *Proc. IEEE International Symposium on Information Theory (ISIT 2011)*, Saint Petersburg, Russia, July 31 - August 5 2011.

[26] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 689–699, February 2009.

[27] R. H. Etkin, "New sum-rate upper bound for the two-user Gaussian interference channel," in *Proc. IEEE International Symposium on Information Theory (ISIT 2009)*, pp. 2582–2586, Seoul, Korea, June 28 - July 3 2009.

[28] A. Chaaban and A. Sezgin, "An extended Etkin-type outer bound on the capacity of the Gaussian interference channel," in *Proc. IEEE Asilomar conference on Signals, Systems and Computers (SSC 2011)*, pp. 2582–2586, St. Pacific Grove, CA, USA November 6-9 2011.

[29] J. Nam and G. Caire, “A new outer bound on the capacity region of Gaussian interference channels,” in *Proc. IEEE International Symposium on Information Theory (ISIT 2012)*, pp. 2806–2810, Cambridge, MA, USA, July 1-6 2012.

[30] I. Sason, “On achievable rate regions for the Gaussian interference channel,” *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1345–1356, June 2004.

[31] D. Tuninetti and Y. Weng, “On Gaussian mixed interference channels,” in *Proc. IEEE International Symposium on Information Theory (ISIT 2008)*, Toronto, Canada, July 2008.

[32] I. Sason, “On the corner points of the capacity region of a two-user Gaussian interference channel,” submitted to the *IEEE Transactions on Information Theory*, August 2013, arXiv:1306.4934.

[33] O. Rioul and M. H. M. Costa, “From almost Gaussian to Gaussian: Bounding differences of differential entropies,” *Information Theory and Applications Workshop (ITA 2015)*, UCSD, San Diego, CA, USA, February 1 - 6, 2015.

[34] Y. Wu, S. Shamai (Shitz), and S. Verdú, “Degrees of freedom of the interference channel: a general formula,” in *Proc. IEEE International Symposium on Information Theory (ISIT 2011)*, Saint Petersburg, Russia, July 31 - August 5 2011.

[35] A. Rényi, “On the dimension and entropy of probability distribution,” *Acta Mathematica Hungarica*, vol. 10, no. 1-2, March 1959.

[36] S. N. Diggavi and T. M. Cover, “The worst additive noise under a covariance constraint,” *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3072–3081, November 2001.

[37] I. Shomorony and A. S. Avestimehr, “Worst-case additive noise in wireless networks,” *IEEE Transactions on Information Theory*, vol. 59, no. 6, March 2013.

[38] E. A. Abbe and L. Zheng, “Coding along hermite polynomials for interference channels,” *IEEE Information Theory Workshop (ITW 2009)*, pp. 584 – 588, Taormina, Italy, October 11-16 2009.

[39] B. Bandemer and A. El Gamal, “Communication with disturbance constraints,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4488–4503, August 2014.

[40] R. Bustin and S. Shamai (Shitz), “MMSE of “bad” codes,” *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 733–743, February 2013.

[41] C. Geng, H. Sun, and S. A. Jafar, “On the optimality of treating interference as noise: General message sets,” in *Proc. IEEE International Symposium on Information Theory (ISIT 2014)*, pp. 1777 – 1781, Honolulu, HI, June 30 - July 4, 2014, arXiv:1401.2592.

[42] R. Tandra and A. Sahai, “Is interference like noise when you know its codebook?” in *Proc. IEEE International Symposium on Information Theory (ISIT 2006)*, Seattle, WA, July 9-14 2006.

[43] A. Chaaban and A. Sezgin, “Sub-optimality of treating interference as noise in the cellular uplink,” May 2011.

[44] S. Gherekhloo, C. Di, A. Chaaban, and A. Sezgin, “Sub-optimality of treating interference as noise in the cellular uplink with weak interference,” arXiv:1401.8265, an extended version of the paper presented in the International ITG Workshop on Smart Antennas (WSA), 2012.

[45] F. Baccelli, A. El Gamal, and D. Tse, “Interference networks with point-to-point codes,” in *Proc. IEEE International Symposium on Information Theory (ISIT 2011)*, Saint Petersburg, Russia, July 31 - August 5 2011.

[46] A. S. Motahari and A. K. Khandani, “To decode the interference or to consider it as noise,” *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1274–1283, March 2011.

[47] C. Nair and A. El Gamal, “The capacity region of a class of 3-receiver broadcast channels with degraded message sets,” *IEEE Transactions on Information Theory*, vol. 55, no. 10, p. 44794493, October 2009.

[48] H. F. Chong, M. Motani, H. K. Garg, and H. El Gamal, “On the han-kobayashi region for the interference channel,” *IEEE Transactions on Information Theory*, vol. 54, no. 7, p. 31883195, July 2009.

[49] B. Bandemer, A. El Gamal, and Y.-H. Kim, “Simultaneous nonunique decoding is rate-optimal,” in *Proc. Allerton Conference on Communication, Control and Computing, Monticello, Illinois*, October 1-5 2012.

[50] W. Huleihel and N. Merhav, “Exact random coding error exponents for the two-user interference channel,” submitted to the *IEEE Transactions on Information Theory*, March 2015, arXiv:1503.02389.

[51] S. S. Bidokhti, V. M. Prabhakaran, and S. N. Diggavi, “Is non-unique decoding necessary?” *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2594 – 2610, May 2014.

[52] K. Moshksar, A. Ghasemi, and A. Khandani, “An alternative to decoding interference or treating interference as Gaussian noise,” *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 305 – 322, October 2014.

[53] A. Dytso, N. Devroye, and D. Tuninetti, “On Gaussian interference channels with mixed Gaussian and discrete inputs,” in *Proc. IEEE International Symposium on Information Theory (ISIT 2014)*, pp. 261–265, Honolulu, HI, USA, June 29 - July 4 2014.

[54] ——, “On the two-user interference channel with partial codebook knowledge at one receiver,” 2014, arXiv:1405.1117.

[55] D. Guo, S. Shamai (Shitz), and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[56] M. Peleg, A. Sanderovich, and S. Shamai (Shitz), “On extrinsic information of good codes operating over Gaussian channels,” *European Transactions on Telecommunications*, vol. 18, no. 2, pp. 133–139, 2007.

[57] N. Merhav, D. Guo, and S. Shamai (Shitz), “Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions,” *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1400–1416, March 2010.

[58] R. Bustin, E. Abbe, H. V. Poor, and S. Shamai (Shitz), “Fading additive Gaussian noise channels: A mutual information trade-off,” *in preperation*, 2015.

[59] T. S. Han and S. Verdú, “Approximation theory of output statistics,” *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[60] D. P. Palomar and S. Verdú, “Gradient of mutual information in linear vector Gaussian channels,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141–154, January 2006.

[61] D. Guo, S. Shamai (Shitz), and S. Verdú, “The interplay between information and estimation measures,” *Foundations and Trends in Signal Processing*, vol. 6, no. 4, pp. 243–429, 2013.

[62] S. Shamai (Shitz), “From constrained signaling to network interference alignment via an information-estimation prespective,” *IEEE Information Theory Society Newsletter*, vol. 62, no. 7, pp. 6–24, September 2012.

[63] Y. Wu and S. Verdú, “Functional properties of minimum mean-square error and mutual information,” *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1289–1301, March 2012.

[64] R. Bartle, *The Elements of Integration and Lebesgue Measure*. Wiley Interscience, 1995.

[65] R. Bustin, R. F. Schaefer, H. V. Poor, and S. Shamai (Shitz), “On MMSE properties of codes for the Gaussian broadcast and wiretap channels,” 2015, arXiv:1501.05994v2.

[66] G. Zitkovic, “Uniform integrability,” *course notes: Theorem if Probability II, Fall 2015: Lecture 12*, 2015, http://www.ma.utexas.edu/users/gordanz/notes/.

[67] G. B. Folland, *Real analysis*. New York: John Wiley & Sons Inc., 1999, pure and Applied Mathematics (New York) (Second edition ed.).

[68] J. S. Rosenthal, *A first look at rigorous probability theory (Second edition ed.)*. Hackensack, NJ: World Scientific Publishing Co. Pte. Ltd., 2006.

[69] R. B. Sowers, “Elementary probability theory,” *course notes, 2012*, 2012, http://www.math.uiuc.edu/∼r-sowers/561Notes/561Notes.pdf.

[70] S. Lalley, “Conditional expectation,” *course notes: Statistics 385, Fall 2014/ Brownian Motion and Stochastic Calculus*, 2014, http://galton.uchicago.edu/∼lalley/Courses/385/.

[71] ——, “Conditional expectation and martingales,” *course notes: Statistics 313 Spring 2013/ Stochastic Processes II*, 2013, http://galton.uchicago.edu/$\sim$lalley/Courses/313/index.html.

[72] S. Shamai (Shitz) and S. Verdú, “The empirical distribution of good codes,” *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 836–846, May 1997.

[73] A. DasGupta, *Asymptotic Theory of Statistics and Probability*. Springer text in Satatistics, 2008.

[74] A. Dytso, D. Tuninetti, and N. Devroye, “On discrete alphabets for Gaussian interference channels with a receiver lacking codebook knowledge,” *Information Theory and Applications Workshop (ITA 2014)*, UCSD, San Diego, CA, USA, February 9 - 14, 2014, arXiv:1401.5536.

[75] Y. Polyanskiy and Y. Wu, “Wasserstein continuity of entropy and outer bounds for interference channels,” April 2015, arXiv:1504.04419.

[76] R. Bustin, M. Payaró, D. P. Palomar, and S. Shamai (Shitz), “On MMSE crossing properties and implications in parallel vector Gaussian channels,” *IEEE Transaction on Information Theory*, vol. 59, no. 2, pp. 818–844, February 2013.